

⑬ 日本国特許庁 (JP)

⑪ 特許出願公開

⑫ 公開特許公報 (A)

昭58—40684

⑤ Int. Cl.³
G 06 F 15/38

識別記号

庁内整理番号
6913—5B

④ 公開 昭和58年(1983)3月9日

発明の数 1
審査請求 未請求

(全 40 頁)

⑭ 自然言語間の自動翻訳方式

① 特 願 昭56—138586

② 出 願 昭56(1981)9月4日

⑦ 発 明 者 新田義彦

川崎市多摩区王禅寺1099番地株
式会社日立製作所システム開発
研究所内

⑧ 発 明 者 岡島惇

川崎市多摩区王禅寺1099番地株

式会社日立製作所システム開発
研究所内

② 発 明 者 山野文行

川崎市多摩区王禅寺1099番地株
式会社日立製作所システム開発
研究所内

⑩ 出 願 人 株式会社日立製作所

東京都千代田区丸の内1丁目5
番1号

④ 代 理 人 弁理士 薄田利幸

明 細 書

発明の名称 自然言語間の自動翻訳方式

特許請求の範囲

1. (a)単語又は／及び連語に関する訳語情報及び品詞情報を格納した辞書メモリを検索することにより、自然言語で記述された入力テキスト文の単語又は／及び連語にそれぞれ品詞を付与する第1ステップと、
(b)品詞列の形に変換された入力テキスト文を言語的意味をもつ最小単位である句要素に分割し、各句要素にそれぞれ品詞を付与する第2ステップと、
(c)各句要素及び単語・連語に付与された品詞の連なりと、予め定められた品詞列とを比較することにより、各句要素及び単語・連語に、構文的な役割を表わしている構文的役割子を付与する第3ステップと、
(d)構文的役割子の配列から、文、節、準節を表現するパターンを生成し、入力テキスト文を、これらのパターンの組み合わせで表現される

骨格パターンに変換する第4ステップと、

(e)予め定められた変換規則を適用することによつて、骨格パターンを形成している入力言語固有の本文、節、準節の配列順序を、出力言語固有の順序に変換する第5ステップと、

(f)予め定められた変換規則を適用することによつて、変換後の骨格パターンを形成している本文、節、準節を構成する入力言語固有の構文的役割子の並び順序を、出力言語固有の構文的役割子の並び順序に変換する第6ステップと、

(g)変換後の構文的役割子列のそれぞれに訳語を付与することによつて、出力言語の文を生成する第7ステップと、

を具備することを特徴とする自然言語間の自動翻訳方式。

2. 第1項記載の方式において、第1ステップにより単語又は／及び連語に付与される品詞が複数個ある場合に、類制される品詞列のパターンの価額を予め用意しておき、該パターンを参照

することにより、多品詞を有する単語及び／又は連語の品詞を最終的に1個決定することを特徴とする自然言語間の自動翻訳方式。

3. 第2項記載の方式において、使用されるテーブルは、品詞の並びが禁止される禁制品詞列パターンの他に、この禁制が解除される特別の条件を含み、該禁制品詞列パターンと禁制品列条件の両者から単語及び／又は連語の品詞の決定が行われることを特徴とする自然言語間の自動翻訳方式。

4. 第2項記載の方式において、禁制品詞列パターンは、品詞判定すべき単語の前後の複数個の単語の品詞を予め定めた品詞列から形成されることを特徴とする自然言語間の自動翻訳方式。

5. 第1項記載の方式において、第1ステップにより単語及び／又は連語に付与される品詞が複数個ある場合に、予め品詞間の優先度を定めた優先度テーブルを参照することにより、最終的に単語及び／又は連語の品詞を1個決定することを特徴とする自然言語間の自動翻訳方式。

(b) 原形に変換された単語を見出し語として、単語辞書を検索し、該単語を基にして形成される連語が存在しない場合には単語の訳語及び品詞情報を逐次、パツファメモリエリアに格納するステップと、

(c) 入力テキスト文と一致する連語が存在する場合には連語辞書を検索し、連語の訳語及び品詞情報を上記パツファメモリエリアに格納するステップと、

を具備することを特徴とする自然言語間の自動翻訳方式。

8. 第7項記載の方式において、連語辞書の検索の結果、すでにパツファメモリエリアに格納された単語を含む連語が見出された場合には、該連語の訳語及び品詞情報をパツファメモリエリアに格納すると共に、すでに格納されている単語情報を無効にすることを特徴とする自然言語間の自動翻訳方式。

9. 第7項記載の方式において、入力テキスト文と一致する連語が複数個存在する場合には最も

6. 第1項記載の方式において、第1ステップにより単語及び／又は連語に付与される品詞が複数個ある場合に、禁制される品詞列のパターンの種類を予め登録したテーブルを参照し、多品詞を有する単語及び／又は連語の中で、禁制される品詞を除去し、更に2個以上の品詞が残った場合には、予め品詞間の優先度を定めた優先度テーブルを参照することにより、最終的に単語及び／又は連語の品詞を1個決定することを特徴とする自然言語間の自動翻訳方式。

7. 第1項記載の方式において、辞書メモリは単語辞書メモリと連語辞書メモリを含み、上記単語辞書は、単語の見出し語と単語を基にして形成される連語の有無と、単語の訳語及び品詞情報を含み、連語辞書は、連語の見出し語と、連語の訳語及び品詞情報を含み、辞書検索の処理は、

(a) 入力テキストの単語を切り出し、該単語の語尾が変化している場合はこれを原形に変換するステップと、

長い連語の訳語及び品詞情報を上記パツファメモリエリアに格納することを特徴とする自然言語間の自動翻訳方式。

10. 第1項記載の方式において、辞書メモリは、句要素を形成する既知の品詞列パターンと、該句要素の品詞との関係を示すテーブルを有し、第2ステップは、

(a) 入力テキスト文の品詞列と、上記辞書メモリの品詞列パターンとを比較するステップと、

(b) 上記の比較の結果、両品詞列が一致しない場合は単語又は／及び連語を1つのノードとし、一致する場合には句要素を1つのノードとして、各ノードの品詞情報及び単語・連語と句要素との区別を表わす情報を順番にパツファメモリエリアに格納するステップと、

を具備することを特徴とする自然言語間の自動翻訳方式。

11. 第10項記載の方式において、入力テキスト文と、複数個の品詞列パターンとが一致する場合には、最も長い品詞列のパターンを形成する

句要素を1つのノードとすることを特徴とする自然言語間の自動翻訳方式。

12. 第1項の方式において、辞書メモリは単語辞書と、予め定められた複数の単語よりなる固定連語の辞書と、予め定められた単語列の一部にブランクを有し、該ブランクに適当な単語、句、又は節を挿入することによつて連語が形成される可変連語の辞書とを有し、上記単語辞書及び固定連語辞書の検索は第1ステップで行い、可変連語辞書の検索は第2ステップの後で行うことを特徴とする自然言語間の自動翻訳方式。

13. 第1項記載の方式において、辞書メモリは、動詞の単語を含む前後の単語及び句要素の列の品詞列と、その品詞列をとり得る既知の構文的役割子列との関係を示すテーブルを含み、第3ステップは、

- (a)入力テキスト文の中で動詞を検出するステップと、
- (b)入力テキストの動詞を含む前後の単語及び句要素の列の品詞列と、上記テーブルの品詞

翻訳方式。

15. 第1項記載の方式において、辞書メモリは、節、準節、パターンと、それらのとり得る構文的役割子列との関係を定義したテーブルを含み、第4ステップは、

- (a)入力テキスト文を接続詞をもつ単語及び句成分で区切るステップと、
- (b)区切られた区間における構文的役割子の列と、上記テーブルの構文的役割子列とを比較するステップと、
- (c)比較の結果、両方の構文的役割子列が一致した場合には、その構文的役割子列に対応する節、準節のパターンを上記区間の入力テキストに付与するステップと、

を具備することを特徴とする自然言語間の自動翻訳方式。

16. 第15項記載の方式において、辞書メモリは、節及び準節のパターンの並びと、骨格パターンとの関係を定めた骨格パターンテーブルを含み、第4ステップは、

列とを比較し、一致する場合にテーブルの品詞列に対応する構文的役割子列を上記単語及び句要素の列に付与するステップと、

を具備することを特徴とする自然言語間の自動翻訳方式。

14. 第1項記載の方式において、辞書メモリは、接続詞の単語の前後の単語及び句要素の列のとり得る既知の構文的役割子列と、その接続詞の単語の構文的役割子との関係を示すテーブルを含み、第3ステップは、

- (a)入力テキストの中で接続詞をもつ単語を検出するステップと、
- (b)接続詞をもつ単語の前後の単語及び句要素の列の構文的役割子列と、上記テーブルの構文的役割子列とを比較するステップと、
- (c)比較の結果、両方の構文的役割子列が一致する場合には、テーブルの構文的役割子列に対応する接続詞の単語の構文的役割子を該単語に付与するステップと、

を具備することを特徴とする自然言語間の自動

- (a)入力テキストに付与された節、準節パターンの並びと、上記テーブルの節及び準節パターンの並びとを比較するステップと、

- (b)比較の結果、両方の節及び準節パターンが一致した場合にはそのパターンに対応する骨格パターンを入力テキストに付与するステップと、

を具備することを特徴とする自然言語間の自動翻訳方式。

17. 第1項記載の方式において、辞書メモリは、入力言語による骨格パターンと、これに対応する出力言語による骨格パターンとの関係を定めた文型変換テーブルを含み、第5ステップは、

- (a)入力テキストを形成している骨格パターンと、上記文型変換テーブルの入力言語の骨格パターンとを比較するステップと、

- (b)比較の結果、両パターンが一致した場合にはこのパターンに対応する出力言語の骨格パターンを入力テキストに付与するステップと、

を具備することを特徴とする自然言語間の自動

翻訳方式。

18. 第17項記載の方式において、辞書メモリは、入力言語の構文的役割子列と、これに対応する出力言語の構文的役割子列の関定を定めた変換テーブルを含み、第5ステップは、

(a) 入力テキストに付与された出力言語の骨格パターンを構成する文、節、準節の構文的役割子列と、上記変換テーブルの入力言語の構文的役割子列とを比較するステップと、

(b) 両比較結果が一致したとき上記文、節、準節の構文的役割子列をこれに対応する出力言語の構文的役割子列に変換するステップとを含み、上記構文的役割子を付与されたノードのカテゴリーが句要素又は単語になるまで、上記(a)、(b)のステップを繰り返して実行することを特徴とする自然言語間の自動翻訳方式。

発明の詳細な説明

本発明は自然言語間の自動翻訳方式に係り、特に、英語で表わされた文章を日本語の文章に自動的に翻訳する方式に関する。

はまず一般的に品詞が確定できる単語を選び、確定した品詞の単語を基準にして、その前後の調休より他の単語の品詞を決定する。最後まで品詞が定まらない単語についてはとり得る品詞の候補を登録しておき、とりあえず一つの品詞を仮に決定する。次に句及び節に相当する品詞列パターンを予め登録しておき、このパターンと一致する品詞列を探す。一致するものがあればその句又は節は1個の品詞記号により置き換えられる。

例えば第1図に示すように… a pulse of known rate of rise という文があつた場合、a pulse, known rate, がそれぞれ名詞句(NP)として判定される。さらに, of + rise が形容詞句(AP)、known rate + of + rise が名詞句、of + known rate + of + rise が形容詞句というように置き換えられる。こうして一つの文が簡単な品詞列のパターン変換される。この変換された品詞列のパターンが予め登録された文を表わす標準の品詞列パターンと比較される。両パターンが一致する場合には翻訳可

本発明の方式は、英語-日本語間の翻訳のみならず、任意の異なる自然言語間の翻訳に適用することができると共に、表記法の異なる同系列言語間の翻訳(変換)方式、例えばカナ文字で記載された文章を漢字混じりの文章に変換する場合にも適用できる。以下の説明では便宜上、入力言語が英語、出力言語が日本語の場合について述べるが、本発明方式はこれに限定されるものではない。

従来技術とその問題点

一つの自然言語で表わされた文章を他の自然言語で表わされた文章に自動的に翻訳する方式として、従来、電気通信学会雑誌第46巻、第11号の第1730頁～第1739頁に記載されている方式が知られている。

この従来方式を簡単に説明する。まず英文テキストが入力されると、辞書検索を行い、単語の並びからなる文章を品詞列に変換する。しかしながら多くの場合、品詞が一義的に定まらないことがある。例えば“study”という単語は動詞をとる場合と名詞をとる場合がある。このような場合に

能と判断され、予め定められた規則に従つて単語の語順変換が行われる。

変換された品詞列パターンが登録された標準パターンと一致しない場合には、単語の品詞の決め方が不適切と判断し、不確定の単語の品詞を登録された別の品詞に置き換え、上記と同じ処理が繰り返して行われる。すなわち、単語が多品詞をもつ場合に、一つの品詞を決定するために一度のフィードバック・ループが用いられる。

最後に語順変換された単語列に訳語が付与され英文を翻訳した和文が出力される。

しかしながら、上記のような従来の自動翻訳方式には下記のような値々の問題点がある。

第1の問題点は多品詞語の決定が極めて複雑になり、翻訳の成功率が低くそなりがちということである。前述のように従来方式では一つの単語が多品詞をもつ場合には、一つの品詞を仮に決めて、パターン辞書による構文解析を行い、うまくいかなかった場合には、又もとに戻つて別の品詞をあてはめてみるという処理を繰り返す。しか

しながら、多数の品詞をもつ単語は現実には数多くあり、しかも文章が複雑になると、一つの文についての可能な品詞列の個数は莫大に増える。これらの品詞列について何回も同じ処理をくり返すことは、翻訳速度の低下を招くことはもちろしのこと、間違つた品詞を暫定的に定めた場合でも、単語の品詞列がたまたま辞書に記録されたパターンと一致し、間違つた翻訳処理が行われてしまうこともある。

従つて文章が複雑になる程、これを正しく翻訳できる率、つまり成功率が低下する。

第2の問題点は、仮に単語の品詞が正しく決定されたとしても、句、節を切り出して1つの品詞におき代える処理を行う際、文頭又は文末から順番に句を切出しており、句、節の係り先、修飾の対象となる単語の関係、つまり依存・修飾関係を分析せずに行つているために間違つた翻訳をしやすいということである。例えば第1(3)図に示すような…take a bus in a city という文があつた場合、第1(4)図と同様に構文分析を行うと、

にパターン辞書による構文分析の処理が含まれている。従つてパターン辞書の登録パターンを追加したり修正する場合には、フィードバックループ全体の動作に矛盾を生じないように処理アルゴリズムを手直ししなければならない。通常、自動翻訳における成功率は上記の構文分析用の登録パターンをいかに構築するかにかかるところ大であり、従つて試行錯誤的にこの登録パターンの増加、変更を要するが、その都度処理アルゴリズム全体の手直しをするのは實際上大きな負担になる。

発明の目的

本発明の目的は上述のような従来技術の問題点を解決した新規な自動翻訳方式を提供することにある。

具体的に言うと、自動翻訳の成功率が高く、複雑な文章を正確に翻訳することができ、しかも文型パターンの増補が容易な方式を提供することが本発明の主たる目的である。

発明の概要

上述の目的を達成するために、本発明方式では

“a bus in a city”が1つの名詞句と認識され、この結果、訳語は“街の中のバスに乗る”となつてしまう。これは“in a city”がtakeにかかる副詞句であるにもかかわらず、文末から順番に句を切出していくと形容詞句として認識され“a bus”を修飾すると判定されたために生じた誤訳である。このように文頭又は文末から順に句を切出していく方法をとると句品詞が一義的に定まらない場合が生ずる。また文章が階層的構造をもつ場合が生ずる。また文章が階層的構造をもつ場合、つまり、修飾をする相手先が複雑に入り組んだ文章の場合にはこれを正確に翻訳することができなくなる。

第3の問題点は、従来方式では翻訳すべき文型の増補を行うためには処理装置のプログラム全体を手直しする必要があり、従つて一旦システムができあがつてしまうと文型の増補が極めて困難になることである。すなわち、従来方式においては、多品詞をもつ単語の品詞を決定するために、フィードバックループを用いており、このループの中

第2図に示す処理過程を経て自然言語間の翻訳が行われる。

まず英文テキストを読み込み、入力データバッファメモリにセットする(テキスト入力処理)。次に辞書を検索してテキスト中の変化形をもつ単語、例えば時制変化、複数形変化、比較級変化などを行っている単語を原形に変換する処理を行う(辞書検索)。更に、テキスト中のすべての単語及び連語(イデオム等)の列を、それらに対応する品詞列に変換する(品詞認識)。次に品詞列の形に変換された文を、誤成組の名詞の連なり、助動詞+動詞、冠詞+名詞、前置詞+名詞、形容詞+名詞等の言語的意味をもつ最小単位、つまり句要素に分割する(句構造認識)。そして上記の各句要素に対して、改めて名詞句、形容詞句、副詞句、動詞句、前置詞句などの句品詞を与える。この後、後述の辞書再検索が行われる。次に句品詞列を、主語、中心動詞、直接目的語、補語、動詞修飾語等の構文的役割子列に変換する。更に構文的役割子列の中から単文パターン、節パターン、

及び意味的に閉じた主語、述語関係を有する英文単位(単語パターン)を見出す(英文型パターン認識)。

更に、前置詞句、副詞句等の句要素の係り先、つまりどの名詞句あるいは動詞句を修飾しているかの依存関係を決定する(依存・修飾関係認識)。

次に上記のようにして認識された英文型パターンを、そのパターン毎に予め定められた変換規則を参照して、日本語の骨格パターンに変換する(文型変換)。最後に予め用意された単語、述語の辞書を検索することにより出力言語である日本語の文を生成する(和文生成)。

本発明方式の特徴の第1は、単語及び述語の列を品詞列に変換する品詞列認識の処理が、句構造認識などの構文分析の前段階において完了し、従来方式のようなフィードバックループをもたないことである。換言すれば品詞列認識の処理ステップではすべての単語・述語の品詞をこの段階で確定してしまい、句構造認識等の処理結果をみて、再び品詞認識に処理が戻るようなフィードバック

ループを有しない。従つて各処理ステップにおける処理結果が他の処理ステップに影響を与えないため句構造等の認識のための登録パターンの増補が極めて容易になる。

特徴の第2は、多品詞をもつ単語の品詞を決定するのに、後述のような品詞認識用パターンを予め定め、これを参照しながら決定するように構成されている点である。この品詞認識用パターンを用いることによつて、各単語の品詞の確定が迅速正確に行われ、結果的に翻訳の成功率が向上することになる。

本発明方式の第3の特徴は、品詞列パターンから句、節を切り出して構文分析を行う際、従来の英文法の句、節とは異なる句要素の概念を導入した点である。更に句要素の単位で切出し、新たに句品詞列を形成した後、これを構文的役割子列に変換することも特徴の1つである。このように句要素、構文的役割子の考え方を導入することによつて、複雑な文章も正確に翻訳することが可能となる。

第4の特徴は、句、節などがどの語を修飾するのかを認識する修飾・依存関係認識の処理ステップを含んでいることである。この処理を行うことによつて第1図図のように修飾の相手元が入り組んでいる文章も正確に翻訳することが可能になる。

本発明の上記特徴及びその他の特徴は、以下説明する実施例を参照することによつて、いつそう明確になろう。

実施例

第3図は本発明方式の全体構成を示す。同図から明らかなように本発明方式は、入力装置100、翻訳処理装置200、辞書用メモリ300、作業用メモリ400及び出力装置500より構成される。

入力装置100は英文テキストを入力するためのもので、通常、キーボードが用いられる。もちろん、この他の方法により英文テキストを入力することも可能である。例えば磁気テープ、磁気ディスク等に予め英文テキストを記録しておき、公知の磁気記憶制御回路を通して処理装置200に

入力してもよい。また、ファクシミリ等の通信機器を通して送られてくる英文テキストを、公知の通信回線制御回路を通して処理装置200に入力することもできる。

辞書用メモリ300は単語、述語(2以上の単語よりなるイデオム等の語)等の辞書や、翻訳処理のために用いられる規則を定義したテーブルを予め格納しておくためのものである。具体的にいうと、このメモリ300には、単語辞書301、述語辞書302、品詞認識用パターン303、句要素切出用パターン304、文型認識用パターン305、依存・修飾関係認識用パターン306、文型変換用パターン307、和文生成用パターン308が格納されている。

単語辞書301のメモリエリアの内部構造を第4(A)図及び第4(B)図に示す。1番地からN番地のそれぞれのブロックには単語情報レコードが記憶されている。単語情報レコードは単語の見出し語と、その単語の属性を示す単語情報とを含む。単語情報としては、その単語をもとにしてつくられ

る連語（イデオム等も含む）の個数、連語の先頭語地、その単語がとり得る品詞の個数、品詞名、品詞の細分類、各品詞における単語の属性その単語が多義の意味をもつ場合にはその個数、それぞれの訳語、各訳語の語尾活用、各訳語の付属機能語等が含まれる。

第4図図には“STUDY”という英語に相当する単語情報が例示されている。この単語をもとにしてつくられる連語としてはMAKE A STUDY OF, UNDER STUDY, CASE STUDY, など図に示されている。それらの連語のうち先頭に登録されている連語の語地が(76)である。(第5図図参照)またはSTUDYは動詞(V)と名詞(N)の両方を取り得るので多品詞個数(2)、品詞名1として(V)、品詞名2として(N)が登録される。なお品詞の細分類の欄には、例えば動詞の場合、自動詞と他動詞の区別、また自動詞でもその後に補語をとる自動詞、(例えばget, look, make等)とその後に前置詞をとる自動詞(例えばsend, advertise等)との区別等、

連語の属性等を示す連語情報を含む。連語情報には、その連語のとり得る品詞の個数、その品詞名品詞細分類、その品詞における連語の属性、連語が多義に解釈される場合その個数、各訳語、各訳語における語尾活用、付属機能語等が含まれる。

第5図図には“MAKE A STUDY OF”という固定連語に関する連語情報が例示されている。この連語は“研究をする”という動詞句を形成するので、品詞の個数は(1)、品詞名は(V)と登録される。

第5(C)図は可変連語辞書の内部構成を示すもので、その詳細については後述する。

更に、この連語の多義性の個数の欄には(1)、訳語1の欄には“研究をする”が登録される。語尾活用1の欄には訳語1の“研究をする”の“する”がサ行変格活用して用いられる旨の情報が登録される。付属機能語1の欄には、上記連語が目的語(OBJ)をとる場合に“(OBJ)の研究をする”の訳語となり、“の”の付属機能語が必要になる旨の情報が登録される。

品詞に関する更に細かい情報が書き込まれる。

更にSTUDYが動詞として用いられる場合、その意味が“勉強する”と解釈される場合と、“研究する”と解釈される場合とがあるので、多義語個数の欄には(2)、訳語1の欄には“勉強する”、訳語2の欄には“研究する”と登録される。また訳語1及び2共に“勉強”及び“研究”の使の“する”が、状況に応じてサ行変化して用いられるので、語尾活用1及び2の欄にその旨登録される。更に付属機能語(て、に、を、は)としては、“STUDY”が目的語(OBJ)をとる場合には(OBJ)の後に“を”が入るので、その情報が訳語1及び2に対応する付属機能語1及び2の欄に登録される。

第4図図では図示していないが第2品詞名(N)についても上述と同様の情報が登録される。

連語辞書302(第3図)のメモリエリアの内部構成を第5(A)図及び第5(B)図及び第5(C)図に示す。

連語情報レコードは、連語の見出し語と、その

第6(A)図は辞書メモリ300の中の品詞認識用パターンエリア303に格納されるパターンの一例を示す。この品詞認識用パターンは単語又は連語の並びからなる文を、それぞれの単語又は連語の品詞の並びへと変形する場合に用いられる。単語又は連語が複数の品詞を文法的にとり得る場合、品詞の並びの前後関係から、最もふさわしい一つの品詞を最終的に選び出す。この場合、予め、文法的にとり得ない品詞列のパターン、つまり禁制品詞列パターンを登録しておき、このパターンを用いて多数の品詞から一つの品詞を選択する。第6(A)図には6個の禁制品詞列パターン及び禁制解除条件が示されているに過ぎないが、実際にはこのパターンが数十個以上用意される。また入力される文の特性を考慮してパターンの変更、増加、簡略化などを行つてもよい。

第6(A)図において禁制品詞列パターンの*は、品詞を判定すべき単語の位置を表わす。又、D、Fは*のそれぞれ1つ前、2つ前の単語の品詞を表わし、B、. Bは*のそれぞれ1つ後、2つ

後の単語の品詞を抜く。

項番1の*位置の「-V」の記号は、「動詞(V)以外の品詞」という意味である。従つて項番1の品詞列パターンは、「i. の後に動詞(V)の可能性のある単語が並んだ場合、動詞(V)として判定し、他の品詞の可能性はなしとしてよい」という意味である。

禁制解除条件は禁制品詞列パターンの禁制が解除されること、換言すれば品詞列パターンの品詞の並びが許容されることを示す。例えば、項番2は「形容詞(ADJ)+動詞(V)」という品詞の並びは通常は禁止されるが、F₁として冠詞(ART)や、BE動詞(BE)がきた場合にはそのような品詞の並びも許容されることを示している。換言すれば「冠詞(ART)ではなく、かつBE動詞(BE)でないもの+形容詞(ADJ)+*」という品詞の並びがあるとき、*の位置が動詞(V)の可能性はないと判断して良いという意味である。

項番3は、「F₁(動詞(V))+*」という

品詞の並びが出てきた場合、F₁が-ingのついた動詞(ING)のとき、*がbe動詞(BE)のとき、或いはF₁がhelpという単語であるときを除いて、*の位置は動詞をとり得ないと判断して良いことを表現している。

項番4は「代名詞複数(PRNu)+*」という品詞の並びが出てきたとき*の位置は(動詞+*)の形はとり得ないと判断して良いことを示す。同様に項番5は「名詞複数(Nu)+*」という品詞の並びが出てきたとき、*の位置は(動詞+*)の形はとり得ないと判断して良いことを意味する。更に項番6は、「場所の固有名詞(N₁)+*」という品詞の並びがあつたとき、*の位置は動詞現在(V_p)の形はとり得ないと判断して良いことを示している。

この他のパターンについては説明を省略するが要するに上記のような禁制品詞列パターンがテーブルとして前記メモリエリア303(第3図)に格納されているのである。

第7(A)図は辞書メモリ300の中の句要素抽出

用パターンエリア304に格納されるパターン例を示す。この句要素抽出用パターンは、単語及び連語に付与された品詞の並びからなる品詞列から句要素を抽出し、この句要素に新たに品詞を付与することにより句要素品詞列を生成する際に用いられる。ここで句要素とは第7(A)図に示す品詞の並びをもつ単語の並びを称し、通常の英文法で定められる句の概念とは異なる。なお第7(A)図は句要素の一部を示したに過ぎず、実際には多数用意される。

第7(A)図において、「000」なる記号は、それをはさむ品詞が1個以上有限個出現したとき、その全体を切り出すことを意味する。項番1は「副詞(ADV)1個」または「副詞(ADV)+…+副詞(ADV)」という品詞列があつた場合、この列全体を句要素として切り出し、新たに句要素品詞として副詞が付与されることを示している。項番2は、「前置詞(PREP)+名詞(N)」あるいは「前置詞(PREP)+名詞(N)+…+名詞(N)」という品詞列には前置詞的(PRENAL)

の句要素品詞が付与されることを示す。項番3は「助動詞(AUX)+動詞(V)」または、「助動詞(AUX)+…+助動詞(AUX)+動詞(V)」という品詞列があつた場合、この列全体を句要素として切り出し、新たに句要素品詞として動詞(V)が付与されることを示す。項番4、5、6についても同様である。

辞書メモリ300の文法認識用パターンメモリエリア305には、4つのテーブルが格納されている。すなわち動詞パターンテーブル、接続詞パターンテーブル、断パターンテーブル、文パターンテーブルである。

第8(A)図は動詞パターンテーブルの一例を示し、動詞のタイプ番号を手がかりとして動詞をもつノードの前後のノードの構文的役割子を決定するために用いられる。動詞のタイプ番号は単語辞書および連語辞書を検索することにより得られる。動詞のタイプ番号と動詞の例、動詞が前後の名詞相当句を支配するパターンを表1(A)に示す。

表1(A)

記号	品詞名と説明	対応するノードに付与されるカテゴリ
PRN	代名詞	W
N	名詞・複合名詞	W, P
ADJ	形容詞・複合副詞	W, P
CNJ	接続詞・準接続詞	W
V	動詞・複合動詞(下記3項目の形以外のもの)	W
TOV	動詞・複合動詞(T, 不定詞型のもの)	W
ENV	動詞・複合動詞(過去分詞型のもの)	W
INGV	動詞・複合動詞(進行型のもの)	W
AUX	助動詞	W
ADV	副詞・複合副詞	W, P
PREP	前置詞・複合前置詞	W
PRENAL	前置詞的(前置詞+(複合)名詞の型のもの)	P, Q
ART	冠詞	W
INT	間投詞	W
NOT	否定詞	W
COM	コンマ	W
PRD	ピリオド	W
.....

ては後で詳述する。

以上述べたように、辞書メモリ300には、単語、連語辞書の他、各種の規則パターンが予め格納されるが、このメモリ300への情報の書き込み及び読み出しは、書き込み読み出し制御回路201によつて行われる。

一方、作業用メモリ400は、第2図に示した各処理の過程で必要になるテーブルを格納するためのもので、読み出した単語、連語情報を一時的に格納するバッファテーブルエリア401、入力テキストストリームテーブルを格納するエリア402、ワードストリームテーブルを格納するエリア403、ノードストリームテーブルを格納するエリア404、句要素列を格納するエリア405、英文型ノード列を格納するエリア406、和文型ノード列を格納するエリア407、出力テキストを格納するエリア408を含んでいる。尚各テーブルの詳細については後述する。

上記作業用メモリ400の情報の書き込み読み出しは、翻訳処理装置200からの指令に基づき、書

第8図は接続詞パターンテーブルの一例を示し、接続詞を有するノードの前後のノードの構文的役割子列パターンから接続詞のノードの構文的役割子を決定するために用いられる。

第8図は副パターンテーブルの一例を示し、接続詞、句読点等で区切られるノード列の構文的役割子パターンから副パターンを決定するために用いられる。

第8図は文パターンテーブルの一例を示し、副パターンの並びから文パターン(骨格パターン)を決定するために用いられる。

それぞれのパターンテーブルの詳細については後述する。特に第8(A)図中、第8(B)図中、第8(C)図中に出現する記号については、後で第20図等を用いて述べる。辞書メモリ300には、この他に形容詞、副詞の単語、句等がどこにかかるとかを決定するための依存・修飾関係認識用パターン306、文型の変換用に用いられるパターン307、和文生成の際に用いられるパターン308をそれぞれ格納するエリアを有する。各パターンについ

込み読み出し制御回路202を介して行われる。

翻訳処理装置200は、後で詳しく述べるように第2図に示した処理を行うためのプログラムを格納しており、辞書メモリ300の情報を適宜用い、メモリ400をワーキングエリアとして用いて、入力英文テキストを和文に翻訳して出力する。

出力装置500としては和文出力をプリント又は表示するためのプリンタ又はディスプレイ装置等が用いられる。

なお、前述の辞書メモリ300として例えばディスクメモリ、作業用メモリ400としてコアメモリが用いられる。もちろん本発明方式ではこれらのメモリの種類は特定のものに限定されない。又これらメモリを処理装置と別に設けてもよいし、処理装置の中に設けても差支えない。

以下本発明方式による処理手順について詳細に説明する。

(1) テキスト入力処理

テキスト入力処理というのは、入力装置100

(第2図)より入力した英文テキストを、作業用メモリ内の入力テキストストリームエリア402に順次セットする処理をいう。英文テキストの一例を第9図に示す。キーボード等の入力装置100により英文を入力すると、アルファベットに対応するコード番号に変換された後、上記入力テキストストリームエリア402に格納される。

(2) 辞書検索

テキスト入力処理が完了すると、次に、いわゆる辞書引きの作業が行われる。辞書引きとは入力テキストストリームエリア402に格納された、入力テキストの単語の出現の順に、単語辞書301を検索し、該当する単語の単語情報を作業用メモリ400の単語、連語用バッファメモリエリア401に順次格納する処理をいう。例えば第9図のテキストの場合、第10図に示すようにまずHEを見出し語とする単語情報がメモリエリア401のWB(1)番地に格納される。次にWRITEは不規則変化動詞であるので予め単語辞書301に登録されており、その単語情報には"WRITE"

へのポイント(WRITEへの先頭番地)と不規則変化情報(wroteはwriteの過去形)が格納されている。従つてWRITEの単語を辞書引きするとWRITEを見出し語とする単語情報が読出され、この単語情報に不規則変化情報を付加した後メモリエリア401のWB(2)番地に格納される。以後、順にENGLISH、VERY…の辞書引きが行われる。

このように単語辞書の内容と同じ情報を一旦、単語、連語用バッファメモリエリア401にとり込むのは、作業用メモリ400として通常、高速アクセス可能なメモリ、例えばコアメモリが用いられるので、後の品詞認定、依存修飾関係認識、和文生成などの処理過程における辞書検索が高速に行えるためである。従つて、上記の単語・連語バッファメモリ401を使用せずに各処理過程において直接単語辞書301、連語辞書302を検索することもできる。

上述のいわゆる辞書引きの段階で、単語の語尾変化処理及び連語の検出処理が行われる。以下そ

れらの処理の内容を詳しく説明する。

語尾変化処理とは、語尾が活用変化している単語、例えば時制変化、数形変化、比較級変化などを行っている単語を、その原形に変換し、単語辞書301、連語辞書302を検索した後、検索した内容に語尾変化情報を付加した上で単語・連語バッファメモリエリア401にセットする処理をいう。この処理を行うことにより後の品詞認識処理における単語辞書あるいは連語辞書の検索を、単語の原形の見出し語を参照して行うことが可能になる。従つて辞書には、このように語尾変化した単語情報を予め登録しておく必要がなくなり、辞書メモリ容量の低減に効果がある。

説明の便宜上、以下"STUDIES"という複数名詞変化形あるいは三人称単数動詞変化形の単語を例にとつて、原形への変換手順を述べる。他の変化形単語の原形変換処理も同様の考え方で処理できる。

処理対象となる単語は単語を構成する一文字ずつにL、.、L、.、L、…なる符号が与えられる。

たとえば"STUDIES"に対しては、第11(1)図のように末尾のSにL、が、次のDにはL、が、さらに次のIにはL、というように付与される。また第11(A)図のステップ1011に示されているように「L、L、削除」という処理を受けると単語の各文字に付与される符号は第11(2)図のように変化する。すなわち末尾の2文字を削除して再び符号のつけ変えが行われる。

まずステップ1002において辞書検索が行われ、辞書にあれば処理は終了し(1003)、なければステップ1004に進む。ステップ1004、1005、1006、1007、1008はそれぞれ語尾変化が生じているかどうかを判定するための処理である。この例の場合はステップ1008で判定結果がYESとなり、ステップ1010に進む。ここでSTUDIESの末尾のSが削除され、STUDIEとなり、ステップ1016で再び辞書検索が行われる。単純な複数形の場合はこの段階で辞書に登録されていることが多い。(例えばlikes)本例の場合は更にステップ1017に進み末尾がEかどうか判定され、

この結果YESとなるので、更にステップ1018に進む。更にステップ1022に至り、L、L、の文字、つまりIEが削除されSTUDになる。更にステップ1023においてYが付加されSTUDYという形に変換された後、ステップ1024で再び辞書引きが行われる。ここでも、もし辞書になければ未定単語と判断される(1025)。

以上述べたような処理を経て、STUDYの単語が辞書にあると、その単語情報が読み出され、単語情報に語尾変化情報(複数名詞変化形並びに三人称単数動詞変化形)を付加したものが前述の単語・連語パツファメモリエリア401に書き込まれる。

次に単語の並びからなる英文テキストの中の単語を検出する処理について第12図を参照して説明する。連語とは、2以上の単語の結合により意味のある語を形成するもので、イデオムもこれに含まれる。

ここでは一例として“MAKE A STUDY OF”なる連語が検出され、単語・連語パツファメモリ

ものと一致するのは“MAKE A STUDY OF”であることが判る(ステップ1049)。なお、台致する連語が存在しないと判断された場合には、“STUDY”を単なる単語とみて、前述と同様に“STUDY”の単語情報をパツファメモリエリア401に格納することになる(ステップ1045)。また入力テキストが2以上の連語と一致することが検知された場合には、最も長い連語と一致したものを採用する。例えば“MAKE A STUDY”と“MAKE A STUDY OF”の両者と一致した場合には後者が採用される。

さて、現在の例では入力テキストと“MAKE A STUDY OF”が一致した訳であるが、この場合にはまず単語、連語パツファエリア401内にすでに格納されている“MAKE”及び“A”の単語情報に対して無効マークの付与あるいは単語情報の消去が行われる(ステップ1050)。この後“MAKE A STUDY OF”に対応する連語情報が単語・連語パツファエリア401に格納される(ステップ1045)。

特開昭58-40684(11)

エリア401に格納される手順について説明する。まずステップ1040で単語の切出しが行われ、ステップ1041で前述の語尾変化処理が行われる。次にステップ1042で単語辞書検索がなされる。“MAKE”及び“A”については前述と同様に単語辞書301の該当見出し語を検索して、その単語情報を単語・連語パツファメモリエリア401に順次格納していく。ステップ1043では、単語が辞書に登録されていない未知単語かどうか判定する。登録されていない場合にはステップ1047で固有名詞として判断する。次に“A”の次の“STUDY”を見出し語として単語辞書301を検索すると、第4図図に示すように、“STUDY”を含む連語が6個存在することが判る(ステップ1044)。そこで直ちに対応連語部の先頭番地(76)から連語の見出し語の検索が開始される(ステップ1048)。第5図図に示すように、“STUDY”を含む連語としては、“MAKE A STUDY OF”“UNDER STUDY”等が存在することが検知される。このうち入力テキスト中の

上記のような処理により、最終的には入力英文テキスト中に出現する単語及び連語は、すべてその出現順にパツファエリア401に格納されることになる。

なお、上記の例では簡単化のため、未知単語はすべて固有名詞として認識する場合を述べたが、より正確な品詞を決定することもできる。この未知単語の正確な品詞の判定は、前述した「漢制テーブル検索による品詞認識処理」を再度実行することにより、全く同様に実行できる。

(3) 品詞認識処理

品詞認識処理とは単語・連語用パツファ401に格納された各単語及び連語に1つずつ品詞を付与する処理を称し、第13図に示すフローチャートにしたがって処理が行われる。

まずステップ1060において、単語・連語用パツファメモリ401内をWB(1)、WB(2)、WB(3)、…の如く順次走査しながら、単語及び連語の品詞名、品詞細分類、属性情報を読み出し、ワードストリームテーブルエリア403に順次格納してい

V ₁	GOV+OBJ	study
V ₂	GOV+△<THAT>+... GOV+<WH>+<TUV> GOV (+<WH>)+△...	suppose, know, say, tell, snow
V ₃	GOV+IOBJ+ {<WH>+<TUV>} △ (<WH>)+...	snow, tell
V ₁₀	GOV+IOBJ+DOBJ GOV+OBJ+ {COMP} <N>	make
V ₁₁	GOV+OBJ+ {COMP} <ADJ>	paint, regard
V ₁₂	GOV+ {OBJ} +<PREP>+ {<N>} <IT> <INJ>+<N> ...	read, congratulate bring
V ₁₃	GOV+OBJ+ {<ADV>} GOV+ADVAL+OBJ	keep, put

V ₁₄	GOV+OBJ (+<NOT>)+TOV	bring, know
V ₁₅	GOV+OBJ+TOV INGV ENV <TV>+<BE>+ {<ADJ>} <N> + ... GOV+<IT>+ {<ADJ>} <N> } △	ful, have, see, gues, think
V ₁₆	GOV+ {<ADJ>} +<TUV> <N> {<TV>+<BE>} + {<ADJ>} <N> }	appear, seem, prove

く。品詞名とその記号の一部を表1(B)に示す。

表1(B)

動詞のタイプ	動詞が中心動詞 (GOV)として出現し た場合の支配パターン の例	…は、次の語に含まれる部分 パターンの後のSUBJは省略 ()内は無くてもよい部分 <>内は単語品詞 △は、その直前でパターンを 示す	動詞(GOV)の例
V ₁	GOV GOV+CUMP; COV=<BE>		be
V ₂	GOV+CUMP; GOV+<BE> {<THERE> } +GOV+CUMP <IT>	GOV+<BE> 以下すべて<BE>	get, look, make, come, become
V ₃	GOV (+<ADVAL>)		rise, walk, fall look
V ₄	GOV+<PREP>		
V ₅	GOV+(<PREP>+<IT>)+△... GOV+(<PREP>)+△...		agree, see, decide
V ₆	GOV+<TOV>		stop, prefer, want, like, study

3540458-10684(12)

第14(A)図及び第14(B)図はワードストリームテーブルの内部構成を示す。同図から利るように各単語・連語はそれぞれワードレコードを含む。各ワードレコードは第14(B)図に示すような情報を含んでいる。一例として単語“WRITE”についてのワードレコードを説明する。単語/連語の区別指示子には単語である旨の情報(単)が書き込まれる。語の識別番号はテキストストリームテーブル(第9図参照)においてその語が何番目に出現するかを表わすので、この場合は(2)が書き込まれる。更に語長の欄は、単語又は連語を構成する文字数を表わし、“WRITE”の場合は(5)が書き込まれる。先頭文字の文字番号の欄はテキストストリームテーブル402(第9図)において、“WRITE”の頭文字“W”の位置する番地(4)が書き込まれる。更に多品詞の個数の欄には動詞(V)と名詞(N)の2つであるので(2)が書き込まれる。その他、品詞の細分類、多義性の個数、各品詞に対応する先頭訳語部(第4図参照)へのポイントがそれぞれの欄に書き込まれる。

の単語、つまり“HE”、“WROTE”と“VERY”、“SLOWLY”の品詞代名詞(PRN)、動詞(V)と副詞(ADV)、副詞(ADV)をそれぞれテスト用バッファにセットする。この状態を第6(B)図に示す。ここで動詞(V)の番字(7)は、動詞の品詞細分類を示し、目的語をとる他動詞であることを表示している。テスト用バッファ(第6(B)図)の*の位置に形容詞(ADJ)をセットした後、第6(A)図に示した禁制品詞列パターンと比較される。この結果、第6(B)図の品詞列パターンは第6(A)図の項番7のパターンと合致し、(一、V、ADJ、一)なるパターンは禁制されることが検知される。従つて“ENGLISH”の単語が、この文章においては形容詞をとり得ないことが認識される。次にテスト用バッファ(第6(B)図)の*の位置に名詞(N)をセットして再び禁制品詞列パターンと比較される。この場合にはいずれの禁制パターンとも合致しないことが最終的に検知され、“ENGLISH”の品詞は名詞であると決定される。

第13図に戻り、ステップ1061においては、WS欄の番地の単語が多品詞かどうか判定する。周知のように各単語及び連語は多品詞をもつことが多く、従つて、この場合には一つの最も適切な品詞を選ぶ処理を行わなければならない(以下この処理を多品詞解消処理と称する)。この多品詞解消処理を実行するために本発明方式では辞書メモリ300に格納された品詞認識用パターン表(第6(A)図参照)が用いられる。単語又は連語が多品詞をもつ場合、ステップ1062において、品詞名の1つがテスト用バッファメモリ(作業用メモリの一部が用いられるがここでは図示しない)にセットされる。ステップ1063においてはテスト用バッファメモリにセットされた品詞列と、第6(A)図に示した禁制品詞列パターンとが順次比較され、一致するパターンを探す。第9図に示す入力テキストの“ENGLISH”の多品詞解消処理を行う場合を例にとつて説明する。“ENGLISH”は名詞(N)及び形容詞(ADJ)の2つの品詞を有する。まず“ENGLISH”の前接の2つずつ

再び第13図に戻り、ステップ1064において、更に第6(A)図に示した禁制解除条件1、2を満足しているかどうか判定される。

こうして禁制品詞列パターンに合致し、且つ解除条件も満足しない場合には適切な品詞の候補から消去される(ステップ1065)。品詞の候補が2個以上残っている場合はステップ1062に戻り同様の処理が繰り返される。この処理によつて最後に1個だけの品詞が残つたかどうかを判定し(ステップ1067)、YESの場合には残つた品詞をWS欄番地の単語又は連語の品詞と決定する(ステップ1069)。

2個以上の品詞が最終的に残つた場合にはステップ1070に進み、ここで第15図のような優先度表による品詞決定が行われる。2つの品詞、例えば形容詞(ADJ)と名詞(N)とが残つてしまつた場合、ADJを見出し語とする第3行と、Nを見出しとする第2列とを参照することによりADJ>Nなる関係、つまりADJの方がNよりも優先度が高いということになり結局ADJが採

用されることになる。

以上、多品詞解消処理の一例について述べたがこの処理の効率、成功率を高めるために下記のような処理をすることもできる。

- (1) テスト用バッファメモリに候補となる品詞をセットする際、出現順度の高い順にセットすることにより処理効率を上げることができる。
- (2) 第15図に示した優先度表は、2個以上の品詞が最終的に残った場合に使用するものとして説明したが、品詞の候補が最終的に全部消去されてしまった場合にも、もとの品詞候補全体に対して第15図の表を用い、最も高い優先度の品詞を採用することができる。

(4) 句構造認識

句構造認識とは、英文入力テキストの各単語及び連語に唯一の品詞を与えることによつてつくられた品詞列から句要素に相当する部分を切り出す処理、および、それらに新たな句要素品詞を付与することにより、“品詞付けされた句要素列(句要素品詞列)”を生成する処理をいう。ここで句

要素とは前述のように、英文法の句の概念とは異なり、言語的意味をもつ最小単位の単語、又は/及び連語の組み合わせよりなるものである。例えば名詞+名詞、助動詞+動詞、冠詞+名詞、形容詞+名詞、前置詞+名詞等は句要素を形成する。

つまり、従来の英文法では、“句”という概念は、かなりゆるやかな定義機能しか持つておらず、与えられた英文テキストに対して、どの部分が“句”かについて一意に定めることができない。例えば本発明でいう“中心動詞となつている句要素”と、“その目的となつている句要素”は、従来の英文法ではそれぞれ単独でも“句”であるが、同時にまたそれらを連結したものの“句”である。英文例をあげると“助動詞+動詞+冠詞+形容詞+名詞”(will have a beautiful girl)は、英文法の“句(動詞句)”とみなされるが、本発明でいう“句要素”ではない。“助動詞+動詞”(will have)と“冠詞+形容詞+名詞”(a beautiful girl)とが“句要素”である。

さらに多くの句を連結させた句、互いに重複し

た部分をもつ異なる句なども、従来の英文法では定義され得る。

これに対し本発明では、中心動詞と、その目的語とは、あくまでも別個の“句要素”である。本発明でいう“句要素”は一般に、与えられた英文テキストに対して、一意に、ただ一つ、互いに重複することなく定まるものであり、これらの組み合わせを、文型認識ならびに文型変換の基礎データとしている。

このことにより、句構造認識の処理を簡単かつ明確にし、かつ、それ以後の文型変換や和文生成等の処理を明確に分離することができる。

第16図は句構造認識の処理の流れを示す。まずステップ1080において、ワードストリームメモリエリア403の各単語及び連語のワードレコードが、ノードストリームメモリエリア404にセットされる。第17図はセットされた状態を模式図として示したもので*1の行に示したNS(1)、NS(2)……NS(20)はノード番号である。又*2の行には対応する単語が格納されているが、実際

にはワードストリームテーブル(第14(A)図、第14(B)図参照)へのポイントの情報が格納される。

*3の行にはノードのカテゴリー、つまり単語(W)、句要素(P)、節(C)、準節(Q)、文(S)の区別を要する情報が格納される。*4の行には品詞情報及び品詞細分情報が格納される。

以後、このノードストリームメモリエリア404にセットされた情報をもとにして句要素が切出されていく。ステップ1081において、ノードの最末尾の番号に+1した番号をkにセットする。すなわち、この処理により入力英文テキストは、ピリオドやコンマも含めて(k-1)個の単語及び連語から構成されたことになる。次にステップ1082において、ノード番号NS(1)、NS(2)、……NS(k-1)を句要素列テーブルエリア405にセットする。ステップ1083において、句要素列テーブルにセットされたノード番号のうち、n以上の番号から(k-1)に至るまでの番号をC₁、C₂、C₃……C_mとする。ただし、nは繰返しが行われる直前にn=1に初期化されている。

さらにノード $NS(C_1)$, $NS(C_2)$, ... の品詞列パターンに関して、第7(A)図に示した句要素として切出すべき品詞パターンと合致するものが存在するかどうか判定する。例えば $n=1$ の場合は H から始まる文の品詞列、 PRN, V, N ... と登録パターンとが比較され、 $n=2$ の場合は $WRITE$ から始まる文の品詞列 V, N, ADV , ... と登録パターンとが比較される。品詞列が第7(A)図に示す品詞パターンの2個以上と合致した場合には、より長い品詞パターンと一致するものを句要素として切出す。第17図に示した英文テキストの場合には $n=4$ のとき、つまり $VERY$ から始まる文の品詞列が ADV, ADV, COM , ... となり、第7(A)図の登録パターンの項番1と一致するので“ $VERY SLOWLY$ ”が句要素として認識される。同様にして、“ $TO STUDY$ ”, “ $TO READ AND WRITE$ ”, “ $IN ENGLISH$ ”, “ $AT SCHOOL$ ” もそれぞれ句要素として認識される。さて、ステップ1084においては、登録した句要素の品詞パターンと一致したノードの並び

れた親ノードを子供ノードに置換して句要素列テーブルを修正する処理が行われる。かくして、当初句要素列テーブルには1, 2, 3, 4, ... 19, 20の順にノードの番号が並んでいたものが、新しい句要素の生成により1, 2, 3, 21, 6, 7, 8, 9, 22, 24, 25, 26, 20の番号順に並びかえられる。(第17図参照)

なお、英文テキストが、例えば“ $Do you know \dots$ ”のような場合には“ Do ”と“ $know$ ”が句要素を形成する。従つて必ずしも連続した位置にある複数の単語によつて句要素が形成されるとは限らない。

(5) 連語辞書再検索

句構造認識の処理が終了した後、もう一度連語辞書検索の処理が行われる。連語辞書再検索を行う理由は次の通りである。

英文中に出現する連語、つまり単語の連なりにより単語と同様の意味や作用をなすものには2つのタイプがある。1つは“ $MAKE USE OF$ ”の如く、固定された単語列よりなる連語であり、他

$NS(C_1)$, $NS(C_2)$... をまとめて新たなノード $NS(k)$ をつくりこれを句要素列テーブルにセットする。

新しく生成されたノード $NS(k)$ を親ノードと称する。第17図のテキストの例では $NS(4)$ と $NS(5)$ とをまとめて新たなノード $NS(21)$ をつくる。このノードの句要素の品詞は第7(A)図のテーブルより副詞 (ADV) が付与される。新たに形成された親ノードには新たなノード番号が付与されると同時に、子供ノードの番号も登録される。つまりノード $NS(21)$ のメモリエリアに格納される情報は、このノード (21) がノード $NS(4)$ と $NS(5)$ からつくられたものであること、このノードが句要素であること (P)、句要素の品詞は副詞 (ADV) であること等である。

一方、ステップ1083において、 NO と判定された場合はステップ1086に進み、 $NS(C_1)$ そのものを句要素とみなす。つまり、この場合には新しいノードの生成はしない。

次にステップ1085においては、新しく生成さ

の1つは、“ $TAKE \sim INTO CONSIDERATION$ ”のように、ある一定の性質をもつ単語や句や節を上記～の部分に取り込んでから初めて完成された連語を形成するものである。説明の便宜上、前者を固定型連語、後者を可変型連語と称する。上記～の部分には名詞句あるいは名詞節がとりこまれる。第5(C)図に示すように連語辞書には名詞を抜く品詞記号“ N ”を用いて“ $TAKE /N INTO CONSIDERATION$ ”と記述される。ここで“ $/$ ”の印は通常の単語のつづりと区別するための符号である。

前述のように、固定型連語の検索処理は、辞書検索の段階ですでに済んでいるが、可変型連語の処理を行うためには、句構造認識が終了した後再び連語辞書を検索する必要がある。

第18図は連語辞書再検索の処理手順を示す。まずステップ1090において、入力テキストの英文の単語数 k を作業メモリ400の適当なエリアにセットする。次にステップ1091において、ノード $NS(i)$ (第17図参照) にある単語を見出し

語とする可変型連語が存在するかどうか判定される。ただし、 i は繰り返しが行われる直前に $i=1$ に初期化されている。もし存在しない場合は、ステップ1096, 1097に進み、全部の単語数について終了するまで同じ判定が繰り返される。

入力テキストの英文が例えば“THE TEACHER TAKES HIS GOOD ATTITUDE INTO CONSIDERATION”の場合、“TAKE”を見出し語とする連語辞書検索で連語が存在することが認識される。(TAKE OUT, TAKE IN, TAKE INTO, /N CONSIDERATION等の連語がある。)次にステップ1092において、ノードNS(i)の前後に存在する句要素をつなげると、可変型連語辞書にあるパターン一致するかどうか判定する。上記の例の場合、“HIS GOOD ATTITUDE”は句構造認識で名詞句(N)と認識されているので、連語辞書パターンの“TAKE /N INTO CONSIDERATION”と一致する。(第5図図参照)入力英文テキストが複数個の可変型連語パターンと一致する場合には、最も長い

可変型連語パターンを採用する(ステップ1093)。更にステップ1094では、可変型連語辞書の連語と一致する入力テキストの部分を新しいノード(親ノード)とすると共に、新しく生成された親ノードを子供ノードと置換して句要素テーブルを修正する。つまり、可変型連語パターンと対応する子供ノード群を句要素テーブルから除去し、代わりに新しい親ノードを代入する。新しいノードの生成があつた場合には、そのノードを形成している最後の単語、すなわち上記の例では“CONSIDERATION”の次の句要素から再び連語辞書検索が行われる(ステップ1095, 1096, 1097)。

以上説明した連語辞書再検索の処理により、最終的な句要素列テーブルが作られる。

(6) 英文型パターン認識

英文型パターン認識とは、ノードNS(i)を複数個ずつまとめ、予め定められた英文のパターンに分類する処理をいう。この英文型パターン認識は、各ノードに構文的役割子を付与する段階と、この

構文的役割子の配列から文、節、単語を見出す段階とに分けて考えることができる。ここで構文的役割子とは、句要素テーブル内の各ノードが、文章の内部でどのような役割をもっているのか、つまり主語(SUBJ)、中心動詞(QUV)、目的語(OBJ)等のどれに相当するのかを表わすものである。

次に英文型パターン認識の処理の流れを第19図及び第20(A)図を用いて説明する。

前述の句構造認識処理の結果、句要素列テーブルメモリエリア405には第20図の*11, *12, *13, *14の各行のように情報が格納されている。*1の行エリアにはノード番号の情報が格納される。*12の行エリアには各ノードに対応する単語又は連語が格納されるが、実際にはノードストリームテーブルへのポイントの情報が格納される。*13の行エリアにはノードのカテゴリー、つまり、単語(W)、句要素(P)、節(C)、単語(Q)、文(S)の区別を表わす情報が格納されている。*14の行エリアには、

単語・連語の品詞、あるいは単語、節、文のタイプ番号が格納されている。*15の行エリアには英文型パターン認識の処理の過程で得られる構文的役割子の情報が格納される。

第19図のステップ1100においては、句要素テーブル405内の各ノードの品詞を順次調べ、動詞族(動詞及びTU+動詞など)の品詞をもつものが検出される。第20(A)図に示した例ではノードNS(2), NS(9), NS(22), NS(24)が動詞族である。次にステップ1101では、検出された動詞族ノードの動詞の細分情報(すなわち動詞のタイプ)を見出し語(エントリー)として動詞パターンテーブルを検索することによつて、動詞族ノード及びその前後のノードの構文的役割子を決定する。

動詞パターンテーブルの一例は第8(A)図に示されている。項番1は見出し語の動詞がbe動詞のタイプ(V₁)であり、かつ、その前後に、「名詞(N)または代名詞(PRN)」と「名詞(N)または代名詞(PRN)」がある場合、より詳しく

く換言すればノード列NS(1), NS(2), NS(3)の品詞が順に名詞(N)または代名詞(PRN)、動詞(V), 名詞(N)または代名詞(PRN)の場合、ノードNS(1)には主語(SUBJ)、NS(2)には中心動詞(GOV)、NS(3)には補語(COMP)の構文的役割子がそれぞれ付与されることを示している。また項番3には見出し語が自動詞のタイプ(V)であり、かつ、その前に名詞(N)または代名詞(PRN)のノードがある場合、より詳しく換言すれば、ノードNS(1), NS(2)の品詞が順に「名詞または代名詞」、「動詞」である場合には、ノードNS(1)に主語(SUBJ)、NS(2)に中心動詞(GOV)の構文的役割子を付与すべきことを示している。又項番4は、ノードNS(1), NS(2)の品詞が共に不定詞の形をとる動詞の場合には、NS(1)に不定詞形の中心動詞(TOGOV)、NS(2)に不定詞形の目的語(TO OBJ)の構文的役割子が付与されることを示している。

さて、第20(A)図に示した英文テキストの例で

が行われたかどうか判定し、YESの場合にはステップ1103に進む。このステップ1103では句検索テーブル405内の残りのノードの品詞を順次調べ、接続詞(CNJ)の品詞をもつものを検出する。更にステップ1104において、検出された接続詞族ノードについて接続詞パターンテーブルを検索することにより、接続詞族ノードの構文的役割子を決定する。

接続詞パターンテーブルの一例を第8(B)図に示す。同図の項番1は、接続詞(CNJ)の単語綴りが" BUT "であり、この前後の構文的役割子列パターンが、句読点(DEL), BUT、主語(SUBJ)、中心動詞(GOV)であつた場合、単語BUTには等位接続詞を表わす構文的役割子BUT(単語と同綴りの記号)を付与すべきことを示している。項番2は、接続詞の単語綴りが" IF "であり、この単語の前後の構文的役割子列パターンが、主語(SUBJ)、タイプ8の中心動詞(GOV)、IF、主語(SUBJ)と配列されている場合は、上記単語" IF "に名詞節

は、ノードNS(1), NS(2), NS(3)の品詞がそれぞれ代名詞(PRN)、動詞(V), 名詞(N)であるため、第8(A)図の項番6の品詞列と一致することが検出される。従つてNS(1)に主語(SUBJ)、NS(2)に中心動詞(GOV)、NS(3)に目的語(OBJ)の構文的役割子が付与されメモリエリア*15に格納される。更にノードNS(22), NS(24)の品詞列は項番4のパターンに一致し、ノードNS(22)にTOGOV、NS(24)にTOOBJの構文的役割子が付与される。同様にしてノードNS(8), NS(9)にも構文的役割子が付与される。なお、第8(A)図はごく一部の動詞パターンを示したに過ぎず、実際には多数のパターンが用意される。

このようにして、第19図のステップ1101の処理の結果、動詞の前後のノードNS(1), NS(2), NS(3), NS(8), NS(9), NS(22), NS(24)の構文的役割子が決定し、その他のノードは未決定のままである。ステップ1102において、すべての動詞族ノードについて動詞パターン表の検索

を導びくIFを表わす構文的役割子NIFが付与されることを示す。同様に項番3, 4, 5, 6には副詞節を導びくIFを表わす構文的役割子ADIF及び名詞節を導びくTHATを表わす構文的役割子NTHATが付与される規則が示されている。

さて第20(A)図の英文テキストの場合、ノードNS(7)の単語の前後には、接続詞テーブル(第8(B)図)の項番1のパターンと一致する構文的役割子列パターンが存在する。従つて上述のステップ1104の処理を実行することによつてノードNS(7)には、構文的役割子BUTが付与される。なお、ノードNS(6)と、NS(20)についてはそれぞれ品詞(CUMとPRD)をみて、自動的に句読点を表わす構文的役割子(DEL)が付与される。

接続詞をもつすべてのノードについて接続詞テーブルの検索が終了したかどうかを判定し、(ステップ1105)その結果、終了した場合にはステップ1106に進む。このステップ1106では、今迄の処理で構文的役割子が付与されなかつたノー

ド、つまり副詞 (ADV) 及び前置詞的 (PRENAL) の句品詞を有するノードに、構文的役割子が未だ定まらないことを表わす未確定修飾子 (PENDM) が付与される。第20図に示す英文テキストの例では、ノードNS(21)、NS(25)、NS(26)にそれぞれ未確定修飾子 (PENDM) が付与される。

次にステップ1107において、接続詞族ノード、(たとえばAND, BUT, OR, ADIF, HTHAT等の構文的役割子をもつノード)により区切られた構文的役割子列パターンの中に、節パターン及び準節パターンがあるかどうかを検索し、存在する場合には節及び準節パターンテーブル(以下単に節パターンテーブルという)を参照し、その節又は準節に「ノードカテゴリー」、「節又は準節の構文的役割子」及び「節又は準節のタイプ」を付与する。

節パターンテーブルの一例を第8(C)図に示す。同図の項番1は、接続詞又は句読点により区分された構文的役割子列パターンの前に位置する接続

詞(以下先行接続詞という)の構文的役割子が「AND, OR, BUT 又は接続詞が無い(φ)」ときで、且つ上記構文的役割子列パターンが「主語 (SUBJ) + 中心動詞 (GOV)」の場合には、このパターンをまとめて新たな親ノードを形成し、この親ノードのカテゴリーとして「文 (S)」、親ノードの構文的役割子として文章が完成したことを表わす「SENT」、文のタイプとしては中心動詞 (GOV) のタイプ番号 (φ) たとえば「TYP1」を付与すべきことを示している。節、文のタイプは、その節や文を支配している動詞 (中心動詞 (GOV)) のタイプと同じく定められる。動詞のタイプ番号と例については表1(9)に示した。

同様に項番2は、親ノードのカテゴリーが「文 (S)」構文的役割子は文が完成したことを表わす「SENT」そして、文のタイプとしては、中心動詞 (GOV) のタイプ番号 (φ) たとえば「TYP2」なる場合の、子ノードの構文的役割子列を表わす。

項番4は、接続詞又は句読点までの構文的役割子列のパターンが「TO不定詞型の中心動詞 (TOGOV)」+「TO不定詞型の目的語」の場合であつて、そのパターンに先行するノードの構文的役割子が「中心動詞 (GOV)」又は「ING型の中心動詞 (INGGOV)」の場合は、そのパターンをまとめて親ノードとし、その親ノードのカテゴリーを「準節 (Q)」、構文的役割子を「目的語 (OBJ)」、準節のタイプとしては中心動詞 (TOGOV) のタイプ番号 (φ)、たとえば「TYP6」を付与すべきことを示している。

項番6は、区切られたノード列の構文的役割子列のパターンが「主語 (SUBJ) + 中心動詞 (GOV) + 補語 (COMP)」であつて、そのパターンに先行するノードの構文的役割子が「ADIF (副詞節を選びくIF)」である場合、このパターンをまとめて親ノードを形成し、その親ノードにカテゴリーとして「節 (C)」、構文的役割子として「CADV (clausal adverb)」、節のタイプとしては中心動詞 (GOV) のタイプ

番号 (φ)、たとえば「TYP2」を付与すべきことを示している。

再び第19図の説明に戻り、ステップ1107で行われる処理を具体的に述べる。第20図に示す英文テキストの場合、ノードNS(1)、NS(2)、NS(3)の構文的役割子列のパターンは、第8(C)図の項番3のパターンと一致する。従つてこれらのノードNS(1)、NS(2)、NS(3)をまとめて新しい親ノードNS(27)が形成され、*11のエリアに格納される。又*13のエリアにはカテゴリーが文であることを示す情報が格納され、*14のエリアには文のタイプTYP7が、*15のエリアには親ノードは文が完成していることを表わす構文的役割子SENTが格納される。

同様にノードNS(22)、NS(24)の構文的役割子列のパターンは第8(C)図の項番4のパターンと一致し、新たに親ノードNS(28)が作られる。この親ノードNS(28)には、カテゴリーとして準節、準節のタイプはTYP6、構文的役割子は目的語 (OBJ) が付与される。

さらに、ノードNS(8)、NS(9)、NS(28)の構文的役割子列のパターンが第8(四)図の項番3のパターンと一致することが検出され上述と同様に、新たな親ノードNS(29)がつくられ検索の結果得られた情報がその親ノードNS(29)の各エリアに格納される。

このようにしてステップ1107では、構文的役割子列から文、節又は準節が検出され、節パターンに新たなノード番号を付与する。ここでは詳しく説明しないが、新しい親ノードが生成された場合は前に説明した手順と同様に、句要素列テーブル405の子ノードを消去し、その代りに新しく生成された親ノードを置き代える処理が行われる。この場合親ノードには子ノードへのポインタの情報が格納される。従つてステップ1107の終了した時点において句要素列テーブル405におけるノードの配列は、順にNS(27)、NS(21)、NS(6)、NS(29)、NS(25)、NS(26)、NS(20)となる。

次にステップ1108に進み、節ノードが副詞節

(CADV)あるいは形容詞節(CADJ)となるときにはこれらに未確定修飾子(PENDM)を付与する処理が実行される。第20図に示す英文テキストの場合には副詞節及び形容詞節はないので新たに未確定修飾詞を付与されるノードはない。

ステップ1109では、接続詞等で区切られるすべての構文的役割子列パターンについて検索が行われたか否か判定され、その結果YESの場合には、最後の処理ステップ1110に進む。ここでは、ステップ1107の処理の結果得られた節パターンが、予め定められた骨格パターンと一致するかどうか判定し、一致した場合には新しいノードが生成される。この判定には第8(四)図に示す骨格パターンテーブルが用いられる。同図の記号と第20(四)図のメモリアreaに格納される情報との対応関係は第20(四)図に示されている。

すなわち、第20(四)図は、1つのノードに対応して1つ定まる。第20(四)図は第20(四)図中の1列分(*11~*15を1つずつ縦に並べたもの)の省略化表記法である。

第20(四)図中の*14には、単語、句要素の品詞、または、節、準節、文のタイプ番号が格納される。

第20(四)図中の*15には構文的役割子が格納される。

第20(四)図中の左肩の節字*13にはノードの категория (W, P, Q, S) が格納される。

第20(四)図中の*16の分岐表現は、本ノードに属する子ノードへのポインタを表現している。これは第20(四)図中の*12の部分に格納されている情報に対応している。

第20(四)図中の*17の波線表現は、副詞句・節や形容詞句・節の依存・修飾関係を表現する。すなわち波線型矢印の尾部にあるノードが、矢印の頭部にあるノードを修飾することを表現する。

項番1は、「カテゴリーが節(C)となつているノード」の前接が0の場合(ノードが無い場合)、つまり未確定修飾子を付与したノードを除き、節パターンが1個しかないときは、そのカテゴリー(*13)を文(S)に変換すると共に、テ-

ブルの*15のエリアには文が完成していることを示す構文的役割子(SENT)を格納すべきことを示している。

項番2は、カテゴリーが文(S)となつているノードの次に構文的役割子が句読点(DEL)となつているノード、その次にAND, BUT, OR等を構文的役割子としてもつノード、その次にカテゴリーが文(S)となつているノード、その次に句読点(DEL)を構文的役割子としてもつノードのようなノードのパターンが検出されたときには、これらをまとめて新しいノードを生成し、その新しいノードのカテゴリー(*13)には文(S)、*14のエリアには重文であることを示す情報(COMP) *15のエリアには文章として完成していることを表す構文的役割子SENTを格納すべきことを示している。

さて、第20(四)図に示す英文テキストの場合、未確定修飾子のノードを除くノード列NS(27)、NS(6)、NS(7)、NS(29)、NS(20)のパターンは第8(四)図の項番2のパターンと一致することが

慣出される。従つてこれらのノード列をまとめて新しいノードNS(30)を生成し、その各エリア*13、*14、*15に新しい情報が書き込まれる。

かくして、第20(A)図に示す英文テキストは、未確定修飾子を付与したノードを除くすべてのノードが単一のノードNS(30)にまとめられたことになる。このノードNS(30)は1つの英文型骨格パターンと呼ばれる。

(7) 依存・修飾関係認識

依存・修飾関係認識とは、前述の英文型パターン認識の段階で、構文的役割子として未確定修飾子(PENDM)を付与されたノードが、どのノードの単語、句、節等を修飾しているのかを認識、決定する処理をいう。

第21図は依存・修飾関係認識の処理手順を示す。ステップ1120では、ノードNS(i)が、構文的役割子として未確定修飾子(PENDM)を持つものかどうか判定される。この判定の結果、YESであればステップ1121に進み、依存・修飾関係

の認識処理が実行される。一方、NOの場合には、ステップ1122に進み、すべてのノードについてステップ1120の処理が終了したかどうか判定される。もちろん、これが完了していない場合にはステップ1120にもどり、同様の処理が繰り返される。

第20(A)図に示したテキストを例にとつて説明すると、この英文の場合、ノードNS(21)、NS(25)、NS(26)が、ステップ1120の処理の結果、修飾先未決定と判定される。これらのノードの修飾先は、辞書メモリ300(第3図参照)に格納された依存・修飾関係認識用テーブルを参照して決定される。

表 2

項 番	ノードNS(i)に対する条件					修 飾 指 定	
	構文的 役割子	品 詞	品詞 分類	依存位置 前後関係	構文的 役割子	構文的 役割子	修 飾 先
1	W・P・Q・C	ADC		(1)文頭 or (2)前にCOM がある	ADVAL		番号の最も若いノードで構文的 役割子がSENTなるもの
2	同上	PRENAL		同上	ADVAL		同上
3	W	ADV		品詞Vの直 後	ADVAL		直前のGUV又は*GUVなる 構文的役割子をもつノード

4	P	PRENAL		直前に品詞 がNのノ ードなし	ADVAL	下記条件を全て満足するノード (1)品詞が動詞族なる(V, TOV, INGV, ENV)もの (2)ノード番号がより小さい (3)未だ他のADVALノードによ つては修飾されていないものが あればそれを優先 (4)上記(1), (2), (3)を満足するも のがあればそれらのうち最もノ ード番号が大きいもの	
5	P	PRENAL		直前に品詞 がNのノ ードあり	P STADJ	直前の品詞がNのノード	

表2は依存・修飾関係認識用テーブルの一例を示す。同図の項番1は、ノードNS(i)に関する条件として、カテゴリーが単語(W)、句要素(P)、単語(Q)、節(C)のいずれかであつて、その品詞が副詞(ADV)で且つ、そのノードが文頭にあるか或いはそのノードの前に司統点(CUM)がある場合には、そのノードの構文的役割子として副詞的修飾子(ADVAL)を付与し、且の番号の最も若いノードで構文的役割子がSENTなるノードを修飾すべきことを意味している。項番2, 3, 4, 5についてもそれぞれ表に記載されたように、ノードNS(i)に関する条件と、そのノードNS(i)が修飾する相手先のノード番号及びNS(i)に付与される構文的役割子との関係が予め定められている。

第20(A)図に示した英文テキストのノードNS(25)について考えると、このノードの品詞は前置詞(PRENAL)であり、且つそのノードの直前に名詞(N)のノードがないから表2の項番5の規則が適用される。従つて、修飾先のノードは、

品詞が動詞族で且つノード番号が25より小さい範囲で一番大きいものであるから、結局、ノードNS(24)ということになる。そして上記ノードNS(25)の構文的役割子はADVALとなる。

同様にノードNS(26)についても表2の項番5の規則が適用され、修飾先のノードはNS(22)、構文的役割子はADVALとなる。

ノードNS(21)について適用される規則は、表2では省略されているが同様の考え方に従つて、ノードNS(2)を修飾することが最終的に決定される。第20(A)図において④→⑤の表示は、 α のノードが β のノードを修飾していることを示している。

以上の修飾・依存関係の処理により、ノード間の関係がすべて定まり、ノード列から遊離したノードは無くなる。

(B) 文型変換

文型変換とは、ノードの配列の順番を英文型パターンから和文型パターンに変換することをいう。

第22図は上記文型変換の処理の流れを示す。

まずステップ1130において句要素列テーブルに最終に残っているノード、つまりカテゴリーがS、構文的役割子がSENTなるノードを演出し、そのノード番号をkとする。第20(A)図に示した英文テキストの場合にはノードNS(30)が最終的に生成された単一のノードであり、従つてこのノードがNS(k)と扱われる。

次にステップ1131において、ノードNS(k)の子ノード列の並びを読み出し、そのノード番号列を句要素列テーブルにセットする。この場合、ノードNS(k)は子ノード番号列のセットの直前に消しておく。第20(A)図の英文テキストの場合、ノードNS(30)の子ノード列は、NS(27), NS(6), NS(7), NS(29), NS(20)であるからそれらのノード番号27, 6, 7, 29, 20がこの順番に句要素列テーブルにセットされる。ここでは句要素列テーブルに並べられる番号の個数をm個と仮定する。

更にステップ1132において、ノードNS(k)の子ノード列を作業用メモリ400内の英文型ノード列パツファエリア406にセットする。

ステップ1133においては、辞書メモリ300内の文型変換パターンテーブル307を探索し、上記のパツファエリア406にセットされた子ノード列の構文的役割子の並びと一致する英文型パターンをみつける。上記文型パターンテーブルの一例は表3に示されており、その内容については後述する。

表 3

項 番	英 文 型 パ タ ー ン		変換後の和文型パターン	
	親ノード	子ノードの並び	SENT DEL BUT SENT	SENT BUT SENT
1	SENT			
2	SENT		SUBJ GOV OBJ	SUBJ OBJ GOV
3	カテゴリー = govc	SUBJ GOV TUOBJ		SUBJ TUOBJ GOV
...

構文的役割子は SENT DEL BUT SENT であるから表3の項番1の英文型パターンと一致し、従つてステップ1133の処理により項番1に示された和文型パターンに変換される。

第22図の説明に戻り、ステップ1134においては上記のステップ1133の処理により得られた和文型パターンを、和文ノード列バッファエリア407のノードJN(1)にセットする。第23図は英文型ノード列バッファエリア406の内容が、文型変換されて和文型ノード列バッファエリア407に格納された状態を示している。

次にステップ1135では、和文型ノード列テーブルJN(2)内のノードのうち、自己のノードを修飾しているノードがあるかどうか判定し、その結果“ある”場合は、ステップ1136に、“ない”場合はステップ1137に進む。第20(A)図のテキストの場合、ノードNS(27)及びNS(29)については自己のノード(又は自己のノードの子ノード)を修飾するノードがあるのでステップ1136に進む。

ステップ1134では、一致した英文型パターンに対応する和文型パターンを文型パターンテーブルより読み出し、作業用メモリ400内に設けられた和文ノード列バッファエリア407にセットする。

表3の文型パターンテーブルは、英文型パターンを和文型パターンに変換するときの変換規則を示している。項番1は、英文テキストの親ノードの構文的役割子が文(SENT)であり、その子ノード列がSENT, DEL, BUT, SENTの場合は、和文型パターンにおける構文的役割子の並びはSENT, “,” , BUT, SENTになることを意味している。項番3は、親ノードのカテゴリーが文(S)又は節(C)であつて子ノード列の構文的役割子が、主語(SUB)、中心動詞(GOV)、TOを伴う目的語(TUOBJ)の順に並んでいた場合には、和文型パターンはSUBJ, TUOBJ, GOVの順に構文的役割子が並びかえられたものになることを意味している。

第20(A)の英文テキストの場合、子ノード列の

ステップ1136においては修飾するノードを修飾されているノードの直前に割り込ませて和文型ノード列テーブルJN(2)を書き換える処理が行われる。(第23図)

一方、ステップ1137では、 L をノードNS(30)の和文テーブルポインタにセットする処理が行われ、さらに $L = L + 1$ 、 $I = 1$ とした後、ステップ1138の判定が行われる。このステップ1138では、句要素テーブル内でI番目にあるノードのノード番号を n_i とすると、ノードNS(n_i)のカテゴリーが句要素(P)または語(W)かどうか判定される。

第20図のテキストの場合、ノードNS(30)における子ノード列のうち、1番目($I = 1$)のノードはNS(27)であり、このノードのカテゴリーは文(S)であるからステップ1138の判定結果は“NO”となりステップ1141の処理に進む。

上記ステップ1141ではノードNS(n_i)の子ノード列の並びを読み出し、句要素ノード列テーブル内の n_i と置換する処理が行われる。従つてこ

れにより m も当然増加することになる。そして読み出された $NS(n_i)$ の子ノード列は前述の英文型ノード列パツファエリア406(第23図)にセットされる。第20(A)図のテキストの場合、ノード $NS(27)$ の子ノードは $NS(1)$ 、 $NS(2)$ 、 $NS(3)$ であるから、上記英文型ノード列パツファエリア406のノード $EN(1)$ 、 $EN(2)$ 、 $EN(3)$ に、それらのノードの構文的役割子である $SUBJ$ 、 GOV 、 OBJ がセットされることになる。

この処理の後、再びステップ1133の処理に戻り、再び文型パターンテーブル(表3)の検索が行われる。英文型ノード列パツファエリア406に新たにセットされた $SUBJ$ 、 GOV 、 OBJ の構文的役割子の配列は、表3の項番2の英文パターンと一致するから、これと対応する和文型パターン、つまり $SUBJ$ 、 OBJ 、 GOV に変換され、和文型ノード列パツファエリア407の $JN(2)$ にセットされる。更にステップ1135では前述と同様に $JN(2)$ のエリアにセットされたノードのうち自己のノードを修飾しているものがあるかどうか判定される。

列パツファにセットされ、再び文型パターンテーブルの検索が行われる(ステップ1134)。こうして和文型パターンが見出され、最終的に和文型ノードパツファ407のエリア $JN(3)$ に第23図に示す和文パターンがセットされることになる。

以上述べたようにしてノード $NS(k)$ (第20(A)図のテキストでは $k=30$)の全ての子ノードについて文型変換の処理が終了すると、ステップ1140に進む。ステップ1140では句要素列テーブルの内容をすべて消去し、番号 k をセットしておく。こうしてノード $NS(k)$ の文型変換処理が終了する。

(9) 和文生成

和文生成とは、和文型パターンを形成するノードに訳語を付与することにより、和文を生成する処理をいう。

第24(A)図及び第24(B)図は和文生成の処理の流れを示す。

ステップ1150において、句要素テーブル内に最終的に生成された単一のノード、つまりカテ

この結果、ノード $NS(2)$ 、 GOV を修飾しているノードがあることが分かる。従つて修飾しているノード $NS(21)$ の構文的役割子 $ADVAL$ (先に $PENDM$ から $ADVAL$ に変換されている)が GOV の直前に割り込ませて $JN(2)$ のエリアの構文的役割子列を並びかえる処理が行われる。(ステップ1136)従つて最終的に $JN(2)$ のエリアにセットされる構文的役割子列は $SUBJ$ 、 OBJ 、 $ADVAL$ 、 GOV の順番になる。(第23図)

以上のようにしてノード $NS(30)$ の最初($i=1$)のノード $NS(27)$ の処理が終ると、次の子ノード $NS(6)$ の処理に移る。このノード $NS(6)$ のカテゴリーは語(W)であるから、ステップ1138の判定の結果“YES”となり、 $i=i+1$ とされる。つまり次の子ノード $NS(7)$ の処理に進む。このノード $NS(7)$ のカテゴリーも語(W)であるので $i=i+1$ とされ、次の子ノード $NS(29)$ の処理に進む。ノード $NS(29)$ は構文的役割子が文($SENT$)であるから、その子ノード列 $NS(8)$ 、 $NS(9)$ 、 $NS(28)$ の構文的役割子が英文型ノード

リーが S 、構文的役割子が $SENT$ のノードを検出し、そのノード番号を k とする。次にステップ1151において、ノード $NS(k)$ を構成する英文パターンを文型変換処理することにより得られた和文パターンを格納する和文型パツファテーブルへのポイント l を検出する。そして和文型パツファテーブル $JN(4)$ 内に設定されたノード列を読み出し、そのノード番号列を句要素列テーブルに設定する。第20(A)図のテキストの場合には、文型変換により、 $k=30$ 、ノード $NS(30)$ 中の和文型パツファテーブルへのポイント l は $l=1$ となっている。そして和文型パツファ407のエリア $JN(1)$ にノード NS の番号として(27)、(6)、(7)、(29)、(20)が格納されているので、この番号列が句要素列テーブルにセットされる。説明の便宜上、句要素列テーブルにセットされたノード番号の個数を m 個、そして、セットされたノード番号列を n_1, n_2, \dots, n_m とする。

次にステップ1152においては、句要素列テーブル内で i 番目にあるノード番号を n_i とすると

き、ノードNS(n_i)のカテゴリーはP又はWか否か判定される。第23図のバッファエリアJN(1)の最初($i=1$)のノード番号は27であり、ノードNS(27)のカテゴリーはSENTであるから上記の判定結果は“NO”であり、従つてステップ1154の処理に進む。

ステップ1154においては、ノードNS(n_i)中の和文型バッファテーブルへのポイントの値を m とする。そして和文型ノード列テーブルJN(4)内にあるノード列を読み出し、句要素列テーブル内の n_i と入れ替える。これにより m の値も当然増加する。

すなわち第23図の例では、NS(n_i)=NS(n_i)=NS(27)であり、NS(27)中のポイント m は $m=2$ である。そこでバッファエリアJN(4)つまりJN(2)に格納されたノード列(1), (3), (21)が読み出され、句要素列テーブルにセットされたノード番号(27)と入れ換えられる。この結果句要素列テーブル内のノード番号の配列は、(1), (3), (21), (2), (6), (7), (29)。

を探索し、更にワードレコードに関する情報を格納してある単語・連語メモリバッファメモリエリア401(第10図)を探索することにより、該ノードNS(n_i)の日本語としての訳語を付与する。

例えば第25図の $i=1$ のエリアに格納されているノード番号(1)の訳語を付与する場合には、ポイントによりまずワードストリームテーブル403のエリアWS(1)の情報を探索し、更にWS(1)に格納されているポイントにより、単語・連語バッファメモリ401のエリアWB(1)の情報が探索される。このエリアWB(1)には、ノードNS(1)の単語“HE”の訳語が格納されているのでこの訳語情報を読み出す訳である。他のノードについても各ノードに対応するワードストリームテーブルへのポイントを、更に単語・連語バッファメモリへのポイントをたどることにより訳語を読み出すことができる。

次にステップ1156に進み、訳語に対して、時刻係数表現処理が必要か否か判定される。この時刻係数表現処理とは、動詞句の場合に、文型・肯

特開昭58-40684(24)

(20)に変わる。従つて m も5個から8個に変わることになる。この処理の後、再びステップ1152にもどる。そして新たに配列されたノード列の最初($i=1$)のノードNS(1)のカテゴリーが句要素(P)または語(W)か否か判定される。ノードNS(1)のカテゴリーは語(W)であるから、上記の判定結果は“YES”となり、 $i=i+1$ と置かれ、 $i>m$ かどうか判定され(ステップ1153)した後、再びステップ1152の処理にもどる。

以上述べた処理を繰り返すことによつて、句要素列テーブルには第25図に示すようにノード番号が(1), (3), (21), (2), (6), (7), (9), (26), (25), (24), (22), (20)の順に格納される。これまでの処理により入力英文テキストが日本語固有の語順に変換されたことになる。

次にステップ1155に進み、句要素列テーブル内の i 番目のノード番号を n_i とする。そしてノードNS(n_i)に関する情報を格納してあるワードストリームテーブル403のエリア(第14(A)図)

定・否定・受身・時制によりその訳語が変化することに対応する適切な訳語を得るための処理である。したがつて時刻係数表現が必要か否かの判定は、該当ノードが動詞句に対応するものかどうかによつて判定される。換言すれば、ステップ1156ではノードNS(n_i)の構文的役割子がGOV, TOGOV, TOOBJ等であるか否かを調べ、もしそうであれば、時刻係数表現処理の必要ありと判断する。

動詞句の訳の生成については、その動詞句が、

(a) 助動詞を含む場合

(b) 助動詞を含まない場合

の2つに分けて処理される。この処理の流れを第24(B)図に示す。

まずステップ1170においては、ノードNS(n_i)に属している動詞句を解析して、訳語処理装置200中のレジスタ(図示省略)に4つのフラグQ, N, S, Tを下記のようにセットする。つまり、Qは動詞句の文型、Nは肯定、否定の区分、Sは能動、受動の区分、Tは時制によりそれ

表 4

項 番	文 型 Q	肯 否 N	受 身 S	時 制 T	訳 出 指 示	コ メ ン ト
1	0	0	0	1	運用+た	肯 定
2	1	0	0	1	運用+たのですか?	疑 問
3	0	0	1	1	未然+れた	受 身
4	1	0	1	1	未然+れたのですか?	受身, 疑問
5	0	1	1	1	未然+なかつた	否 定
1	0	0	0	0	終止形	肯 定
2	1	0	0	0	運用+ますか?	疑 問
3	0	1	0	0	未然+ない	否 定
4	1	1	0	0	未然+ないのですか?	疑問, 否定
5	0	1	1	0	未然+れない	受身, 否定

ぞれ0または1の値がセットされる。

次にステップ1171に移り、ノードNS(n_1)の子供ノード連の中に、品詞として助動詞(AUX)を持つものがあるかないか調べる。あればYESとしてステップ1173に移り、無ければNOとしてステップ1172に移る。ここでは、まずNOの場合、つまり上記の(b)の場合について説明する。

ステップ1172では、翻訳処理装置200中のレジスタから4つのフラグQ、N、S、Tの値を読み出し、そして、辞書メモリ300の和文生成用パターンメモリエリア308に格納された「語尾決定用テーブル」と比較照合することにより、活用語尾と付加語尾とを決定する。決定された語尾は翻訳処理装置200中のレジスタに一時記憶しておく。「語尾決定用テーブル」の一部を図4に示す。図4では、Q=0のとき肯定文、Q=1のとき疑問文、N=0のとき肯定形、N=1のとき否定形、S=0のとき能動形、S=1のとき受動形、T=0のとき現在形、T=1のとき過去形を示す。

たとえば、図4の項番1は、文章が肯定文(Q=0)で、動詞句が肯定形(N=0)、能動形(S=0)、かつ、時制が過去形(T=1)の場合、動詞句の訳文は、動詞の運用形の後に“た”を付加すべきことを示している。

以下、具体例について、語尾決定の処理を述べる。

たとえば、ノードNS(n_1)に属する動詞句(ただしアンダーライン部分のみ)が“I STUDIED ENGLISH”である場合、第24(a)図のステップ1155により単語辞書(第4図参照)から訳語“勉強”が取り出され、かつまた語尾活用は“サ変(サ行変格活用)”であるという情報が取り出される。これらの訳語・活用情報は、翻訳処理装置200中のレジスタに一時セットされる。

次にステップ1156において、動詞句であるためYESと判定されてステップ1157へ移る。ステップ1157は、すなわち第24(c)図である。第24(c)図のステップ1170により4つのフラグQ=0(肯定文)、N=0(肯定形)、S=0(能

動形)、T=1(過去形)が翻訳処理装置200中のレジスタにセットされる。

助動詞(AUX)はないので、ステップ1171ではNOと判断され、ステップ1172に移る。

ステップ1172では、第3図中の和文生成用パターンメモリエリア308に格納されている「語尾活用テーブル(図4)」を探索する。すると(Q、N、S、T)のパターンは、項番1と一致するので、項番1の訳出指示の「運用+た」が探索される。一方、既に、翻訳処理装置200中のレジスタには、(訳語、活用)情報として(勉強、サ変)があるため、“サ変の運用形+た”、つまり“した”なる語尾が“勉強”という訳語に付加されて、“勉強した”という語尾付の訳語が決定される。

次にステップ1173に移った場合、つまりノードNS(n_1)の子供ノード連の中に品詞として助動詞(AUX)を持つものがあり、ステップ1171でYESと判断された場合、つまり前述の(a)の場合について説明する。

実施例の一部を表 6 に示す。表 6 の説明は後述する。

次に表 5 の解釈の仕方を説明する。

ステップ1173では、まず、辞書メモリ300の和文生成用パターンメモリエリア308(第3図参照)に格納されている「助動詞細分類テーブル」を比較参照しつつ、ノードNS(n_i)に属している助動詞句中の助動詞の分類コードを決定する。この決定された分類コードは一時、翻訳処理装置200中のレジスタに記憶される。「助動詞細分類テーブル」の一実施例の一部分を表5に示す。表5の説明については後述する。

次にステップ1174に移る。このステップでは既に、ステップ1170で決定されている「(Q, N, S, T)フラグ情報」と、ステップ1173で決定された「助動詞分類コード情報」を読み出し、そして辞書メモリ300の和文生成用パターンメモリエリア308(第3図参照)に格納されている「助動詞を考慮した語尾決定用テーブル」とを比較参照することにより活用語尾と付加語尾を決定する。決定された語尾は、翻訳処理装置200中のレジスタに一時記憶しておく。

「助動詞を考慮した語尾決定用テーブル」の一

列 查		1	2	3	4	5	6	7	8	9	10	11
項	時刻	V(原形)	V(過去)	V(pp)	V(ing)	be	be	have	have	have	been	been
番	番分類					V(ing)	V(pp)	V(pp)	V(ing)	V(pp)	V(ing)	V(pp)
1	do	DO1										
2	can	CA1				CA2	CA3	CA4	CA5	CA6		
3	may	MA1				MA2	MA3	MA4	MA5	MA6		
4	must	MU1				MU2	MU3	MU4	MU5	MU6		
5	shall	SH1				SH2	SH3	SH4	SH5	SH6		
6	will	WI1				WI2	WI3	WI4	WI5	WI6		
7	could	CU1				CO2	CO3	CO4	CO5	CO6		
8	might	MI1				MI2	MI3	MI4	MI5	MI6		
9	should	SD1				SD2	SD3	SD4	SD5	SD6		
10	would	WO1				WO2	WO3	WO4	WO5	WO6		

同表の縦軸には動分頻された助動詞が列挙されており、横軸には時制として原形、過去、過去分詞 (pp)、進行形 (ing) …等が列挙されている。DU1, CA1, CA2, …, MA1, MA2, …等は助動詞分類コードである。

ノードNS(n₁)に属する動詞句を分析して、同義の縦軸と横軸の入り口(エントリー)を見つけ、分類コードを読み出すのである。

たとえば項番9と列番7の交叉する部分については、助動詞句か、「should have + V(pp) (過去分詞型の動詞)」の形で出現している場合には、分類コードとして、SD4が検索決定されることを示している。

次に表 6 の解釈の仕方について説明する。

表 6

列番	1	2	3
項番			
1	助動詞 AUX	$Q=0, N=0, S=0$	$Q=0, N=1, S=0$
2	MA1 (適用)てもよい (終止)かもしれない	てもよいか	ではならない
3	MA2 (適用)ていらいかもしれない		
4	MA3 (適用)れらかもしれない		れないかもしれない
5	MA4 (適用)てしまつたかもしれない		れなかつたかもしれない
10	MU1 (未終)なければならぬ (終止)にらがない	なければならぬか にらがないか	(連体)必要はない (未然)にらがない

11	MU2 (適用)ているにらがない		ていないにらがない
12	MU3 (未然)れなければならぬ	れなければならぬか	
13	MU4 (適用)たにらがない	たにらがないか	(未然)なかつたにらがない
21	SD1 (終止)べきである	べきであるか	べきではない
22	SD2 (適用)ているはずだ		ていないはずだ
23	SD3		
24	SD4 (終止)べきであつた		べきではなかつた

表6の縦軸には、助動詞の分類コードが列挙されており、横軸には、助動詞の肯定、否定、疑問などの区別が前述の(Q, N, S)フラグの1または0の値により列挙されている。ノードNS(n₁)に属する助動詞を解析して得られる(Q, N, S)フラグ値と、助動詞分類コードとを入口(エントリ)として、活用語尾と付加語尾とを選択決定するのである。

以下具体例について助動詞を考慮した語尾決定の例を述べる。たとえばノードNS(n₁)に属する助動詞(ただしアンダーライン部分のみ)が、
 “you should have taken his opinion into consideration”である場合、第24(A)図のステップ1155により、連語辞書(第5(C)図参照)から訳語“/Nを考慮”が取り出され、かつまた語尾活用は“サ変(サ行変格活用)”であるという情報が取り出される。これらの連語・活用情報は、翻訳処理装置200中のレジスタに一時セットされる。次にステップ1156において、助動詞であるためYESと判定され、ステップ1157へ進む。

ステップ1157は、すなわち第24(C)図である。第24(C)図のステップ1170により4つのフラグ、 $Q=0$ (肯定文)、 $N=0$ (肯定形)、 $S=0$ (能動形)、 $T=*$ (セットせず)が、翻訳処理装置200中のレジスタにセットされる。次にステップ1171に移るが、助動詞(AUX)が存在するので、ステップ1171では、YESと判断され、ステップ1173に移る。ステップ1173では、第3図中の和文生成用パターンメモリエリア308に格納されている「助動詞細分類テーブル(表5)」を検索し、項番9と列番7の交叉するところのコード“SD4”を検索し、翻訳処理装置200中のレジスタにセットする。

次にステップ1174に移る。このステップでは、辞書メモリ300の和文生成用パターンメモリエリア308(第3図参照)に格納されている「助動詞を考慮した語尾決定用テーブル(表6)」の比較参照がおこなわれる。既に、ステップ1170により、“ $Q=0, N=0, S=0$ ”なるフラグ値がセットされているため、表6の列番1のエン

トリーが選択される。また、既にステップ1173により、助動詞分類コード“SD4”がセットされているため、表8の項番24のエントリーが選択される。したがって、表6からは語尾情報としては、“(終止)べきであつた”が選択決定される。この決定結果と、既にステップ1155により検索されていた、訳語情報“/Nを考慮(すなわ)”とが、照合処理されて、“/Nを考慮するべきであつた”という語尾付きの訳語が決定される。

なお、/Nについては、“his opinion”に対応するノードに対する和文生成処理により、

彼の意見”という訳語が組み込まれることはもちろんである。

以上説明した時刻・領域検索処理の実行が終了すると、表24(図)のステップ1158の処理に進む。このステップ1158では、訳語に対して付属語(て、に、を、は)を付加する処理が必要か否かを判定して、必要な場合はステップ1159に進む。この判定は該当するノードNS(0_i)の構文

属記号がSUBのときは、付属語として“は”および“が”をとり得ることを示し、DOBのときは、“を”および“に”をとり得ることを示している。

表 7

記号	SUB	DOB	
付属語	は	か	を
			に

付属語記号が単一の付属語をとり得る場合は、付属語がそのまま付属語記号に置き換えられる。上記のように付属語記号が2個以上の付属語をとり得る場合は、次のようにして、単一の付属語が選択される。

まずSUBの場合は、当該ノード(構文的役割子がSENTのもの)が他のノードを参照する場合に“が”が付与され、その他の場合には“は”が付与される。

一方DOBについては、動詞の訳語情報に、

的役割子が、主語族(SUBJ, CSUB, TOSUB等)あるいは目的語族(OBJ, TEOBJ, IOBJ, COBJ等)である場合にYES、そうでない場合にNOとすることにより行われる。ここでSUBJは主語、CSUBは「節の形をした主語」、TOSUBJは「TO不定詞型の主語」をそれぞれ意味し、OBJは目的語、TEOBJは「TO不定詞型の目的語」、IOBJは「間接目的語」、COBJは「節の形をした目的語」をそれぞれ意味する記号である。

ステップ1159における付属語付加処理を表3の項番2を例にとつて説明する。

項番2の和文型パターンは、SUBJ, OBJ, GOVであるが、この場合、まずSUBJの次にSUB、OBJの次にDOBの付属語記号が挿入される。次に、各付属語記号に対応した付属語が割り当てられる。この割り当てに際して、和文生成用テーブル308に格納された付属語テーブルが参照される。

表7は付属語テーブルの一例を示すもので、付

“に”をとり得るというフラグがセットされている(第4(A)図の付属語機能語の欄にセットされている)、“に”を付与し、その他の場合には“を”を付与する。この結果、たとえば“resemble”は“～に似る”と訳され、“study”は“～を勉強する”のように訳される。

このようにして付属語付加処理が終了すると、ステップ1160に進む。ここでは、ノードNS(0_i)に対して最終的に得られた訳語を出力テキストテーブルエリア408に左づめにセットする処理が行われる。その後、ステップ1161に進み1>mの関係を満足するまで、上述の処理が繰り返して実行される。

第26図は和文生成処理の結果、出力テキストテーブルエリア408に得られる訳文を示している。すなわち、*20で示した和文型ノード列エリア407の各ノードに訳語が付与され、出力テキストテーブルエリア408には*21で示す訳語が得られる。

以上、本発明方式の実施例について説明したが、

この方式によれば下記のような効果が得られる。

- (1) 辞書検索の段階で、単語の語尾が時制、複数形などにより変化している場合に、原形にもどした後に辞書引きを行つてゐるので、辞書に用意する単語の数を著しく低減することができる。
 - (2) 単語辞書と連語辞書とが用意され、単語辞書における各単語の情報として、その単語をもとにして形成される連語の有無と、連語辞書へのポイントの情報が含まれているので、単語を切出しながら同時に連語も切出すことができる。
 - (3) テキスト入力後、単語及び連語辞書の内容と同じ情報を一旦、高速アクセス可能なランダムメモリにとり込んでいるので、後の処理における辞書検索を高速に行い得る。
 - (4) 連語辞書として固定連語辞書の他に可変連語辞書を有し、句要素切出しの後に可変連語辞書の検索を行つてゐるので、あらゆるイデオムの検出が、少ない辞書容量で可能になる。
 - (5) 品詞認識処理が、文型パターン認識の処理の前に完了してしまうので、文型パターンの変更
- (9) 英文の文型パターンを認識する場合において、句要素を最下層におき、中間に単語、節ノード、頂点部に文ノードを持つような階層構造を採用しているため、入力言語の文型（英文型）から、出力言語の文型（和文型）への変換が、「単なる兄弟ノード間の縦型並べかえ」という形式で簡略に実現できる。従来のように、入力言語解析木から出力言語解析木への木構造変換（Tree transduce）という複雑な処理を回避することができる。また文型変換規則の変換も容易となる。
- 10 構文的役割子の定まらない句要素に対し、修飾先を定めるための依存修飾関係の認識処理を行つてゐるので、複雑な修飾関係をもつ文章の翻訳も可能になる。又、骨格パターンは、形容詞、副詞のように相手を修飾する構文的役割子をもつものを除いた句要素により構成されるのでその骨格パターンの種類があまり多くならない。
- 11 和文生成の処理に、時制標現表現処理及び付

増補と、品詞認識用規則の変更・増補を独立に設計することができる。

- (6) 単語及び連語の品詞の認識に、禁制品詞列パターンを用い、更に必要に応じて優先度テーブルを用いてゐるので、認識処理を迅速且つ正確に実行できる。
- (7) 従来の英文法における句とは異なる概念の「句要素」を定義し、入力英文テキストを「句要素」の単位で切出し、句要素品詞列を形成した後、骨格パターンをつくつてゐるので、骨格パターンの形成に至る処理及び和文生成の処理が容易になる。従来の英文法における句の概念では、名詞句を含む名詞句、名詞句を含む動詞句などがあり、句の切出しが困難である。
- (8) 英文の骨格パターンを形成するのに構文的役割子の概念を導入し、句要素品詞列を構文的役割子列に変換した後、節、単語を検出し、更に節、単語から文の骨格パターンを形成しているため、複雑な英文に対しても容易且つ正確に、種々のパターン設定、変更が可能である。

属語付加処理を含むので、正確な翻訳が可能になる。

- 12 和文生成の処理において、時制標現処理は、構文的役割子として中心動詞族（UOV族）を持つノードだけ、付属語付加処理は、構文的役割子として主語、目的語族（SUBJ族、OBJ族）を持つノードだけ、のように局域化されているため、これらの処理が簡略化されると同時に修正、変換が容易になる。
- 13 入力英文の解析、文型の変換、および出力和文の生成等の過程における翻訳規則がすべてパターンメモリ内のテーブルの形で記述され、翻訳処理装置は、これらのテーブルを解釈実行するように設計されている。従つて翻訳規則の変更、修正、補強などが、単なるパターンメモリ内部の情報の置換えのみで簡単に、翻訳処理装置とは独立に実行できる。

図面の簡単な説明

第1(A)図、第1(B)図は従来の翻訳方法の説明図、第2図は本発明の自動翻訳方式の概念の説明図、

第3図は本発明の自動翻訳方式の構成図、第4(A)図、第4(B)図は本発明方式に用いられる単語辞書の一例を示す図、第5(A)図、第5(B)図は本発明方式に用いられる固定連語辞書の一例を示す図、第5(C)図は本発明方式に用いられる可変連語辞書の一例を示す図、第6(A)図は本発明方式の品詞認識に用いられる禁制品詞列パターンの一例を示す図、第6(B)図は品詞認識のための説明図、第7(A)図は本発明方式の句構造認識に用いられるパターンの一例を示す図、第8(A)図、第8(B)図、第8(C)図、第8(D)図は本発明方式の英文型パターン認識に用いられるパターンの一例を示す図、第9図は入力テキストストリームテーブルの説明図、第10図は単語・連語バッファメモリの説明図、第11(A)図、第11(B)図、第11(C)図は本発明方式の辞書検索における語尾変化処理の流れを示す図、第11(D)図、第11(E)図は語尾変化処理の説明図、第12図は本発明方式における辞書検索の処理の流れを示す図、第13図は本発明方式における品詞認識処理の流れを示す図、第14(A)図、第14(B)図

図は本発明方式のワードストリームテーブルに格納される情報の内容を説明するための図、第15図は本発明方式における品詞認識処理に用いられる優先度表を示す図、第16図は本発明方式における句構造認識の処理の流れを示す図、第17図は本発明方式のノードストリームテーブルに格納される情報の内容を説明するための図、第18図は本発明方式の可変連語辞書検索の処理の流れを示す図、第19図は本発明方式における英文型パターン認識の処理の流れを示す図、第20(A)図、第20(B)図は句要素列テーブルに格納される情報の内容を説明するための図、第21図は本発明方式における依存・修飾関係認識の処理の流れを示す図、第22図は本発明方式における文道変換の処理の流れを示す図、第23図は英文型ノード列テーブル及び和文型ノード列テーブルに格納される情報の内容を説明するための図、第24(A)図、第24(B)図、第24(C)図は本発明方式における和文生成処理の流れを示す図、第25図は和文生成処理の説明図、第26図は和文生成により、入力

英文テキストに訳文が付与された状態を示す図である。

100…入力装置、200…翻訳処理装置、300…辞書メモリ、400…作業用メモリ、500…出力装置。

代理人 弁理士 海田利率

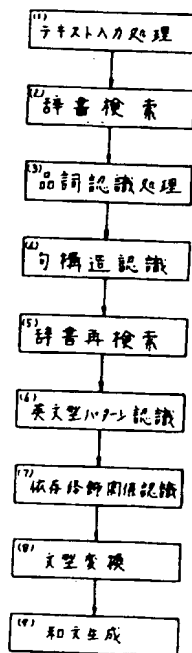
第 1 (A) 図

--- a pulse of known rate of rise
 L NP ———— L NP ———— L NP ————
 L AP ————
 L NP ————
 L AP ————
 L NP ————

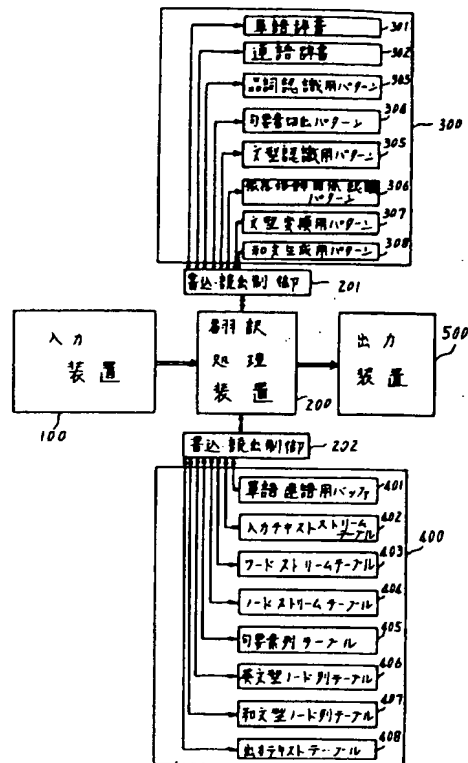
第 1 (B) 図

--- take a bus in a city
 L NP ———— L NP ————
 L AP ————
 L NP ————

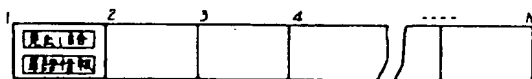
第 2 図



第 3 図



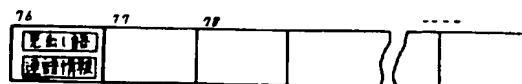
第 4 (A) 図



第 4 (B) 図

見出し語 (STUDY)			
連語の国数 (6)		固定連語部見出し各記 (76)	
多品詞国数 (2)	品詞名1 (V)	品詞名2 (N)	-----
第1品詞名 (V)			
品詞細分類	属性1	属性2	----
英辞細類 (2)	訳語1 (勉強F3)	訳語2 (研究F3)	---
	語尾活用1 (F4"勉強")	語尾活用2 (F4"研究")	---
	付属機能語1 (OBJ"g")	付属機能語2 (OBJ"n")	---
第2品詞名 N			

第 5 (A) 図



第 5 (B) 図

固定連語見出しパターン (MAKE A STUDY OF)			
多品詞国数 (1)	品詞名1 (V)	品詞名2	----
第1品詞細分類	属性1	属性2	----
多義性国数 (1)	訳語1 (研究F3)	訳語2	---
	語尾活用1 (F4"研究")	語尾活用2	---
	付属機能語1 (OBJ"n")	付属機能語2	---
UNDER STUDY			

第 6 (A) 図

第 5 (C) 図

可変型連語見出しパターン (TAKE IN INTO CONSIDERATION)			
多品詞句型 (1)	品詞名1 (V)	品詞名2 (V)	---
第1品詞細分類 (3)	属性1	属性2	---
多品詞句 (1)	語句1 (INを考慮)	語句2	
	語尾活用1 (十動・有動)	語尾活用2	
	付属機能語1 REL	付属機能語2	

項番	禁制品詞列パターン					禁制解除条件1			禁制解除条件2	
	F ₂	F ₁	*	B ₁	B ₂	1	2	3		
1	-	to	-	V	-					
2	-	ADJ	V	-	-	F ₂	ART	F ₂	BE	
3	-	V	V	-	-	F ₁	ING	*	BE	F ₁ HELP
4	-	PRN	V ₅	-	-					
5	-	N _M	V ₅	-	-					
6	-	N ₁	V _P	-	-					
7	-	V ₇	ADJ	-	-					

第 6 (B) 図

PRN	V ₇	ADJ	ADV	ADV
F ₂	F ₁	*	B ₁	B ₂

第 8 (A) 図

第 7 (A) 図

項番	句要素1(12動・有動・品詞パターン)				句要素2(12動・有動・品詞)			
1	ADV	---	ADV					ADV
2	PREP	N	---	N				PRENAL
3	AUX	---	AUX	V				V
4	AUX	NOT	V					V
5	TO	V						TOV
6	V	AND	V					V

項番	句要素1(12動・有動・品詞パターン)	品詞列 (0は禁止語)	句要素2(12動・有動・品詞)	句要素3(12動・有動・品詞)	句要素4(12動・有動・品詞)	句要素5(12動・有動・品詞)	句要素6(12動・有動・品詞)	句要素7(12動・有動・品詞)	句要素8(12動・有動・品詞)
1	V ₁	N PRN *	N PRN	SUBJ	GOV	COMP			be
2	V ₁	N PRN *	N PRN		GOV	COMP			
3	V ₃	N PRN *	N PRN	SUBJ	GOV				
4	TOV ₆	N PRN *	TOV		TOGOV	TOOBJ			like
5	TOV ₆	N PRN *	V	SUBJ	GOV	TOOBJ			
6	V ₇	N PRN *	N PRN	SUBJ	GOV	OBJ			write
7	V ₁₀	N PRN *	N PRN	N SUBJ	GOV	OBJ ₁	OBJ ₂		give
8	V ₁₁	N PRN *	N PRN	ADJ SUBJ	GOV	OBJ	COMP		pain

第 8 (B) 図

特開昭58-40684(33)

項番	元の文の構文	構文的制約子列ハローン (※57は元の文)	構文的制約子
1	BUT	DEL * SUBJ GOV	BUT
2	IF	SUBJ GOV (VB) * SUBJ	NIF
3	IF	* SUBJ GOV	ADIF
4	IF	SUBJ GOV OBJ *	ADIF
5	IF	SUBJ GOV OBJ DEL	ADIF
6	THAT	SUBJ GOV * SUBJ	NTHAT

第 8 (B) 図

第 8 (C) 図

項番	元の文の構文	次の構文制約子は元の文の直前のハローン	構文ハローン
1	φ, AND OR, BUT	SUBJ GOV	TYP* SENT ^S * = 1, 2, 3 (先行詞を含む)
2	φ, AND OR, BUT	SUBJ GOV COMP	TYP* SENT ^S * = 1, 2 (")
3	φ, AND OR, BUT	SUBJ GOV OBJ	TYP* SENT ^S * = 7 (")
4	GOV, IN & GOV	TO GOV TO OBJ	TYP* OBJ ^O * = 6, 8 (")
5	GOV, IN & GOV	TO GOV OBJ	TYP* OBJ ^O * = 7 (")
6	ADIF	SUBJ GOV COMP	TYP* CADV ^C * = 1, 2 (先行詞を含む)

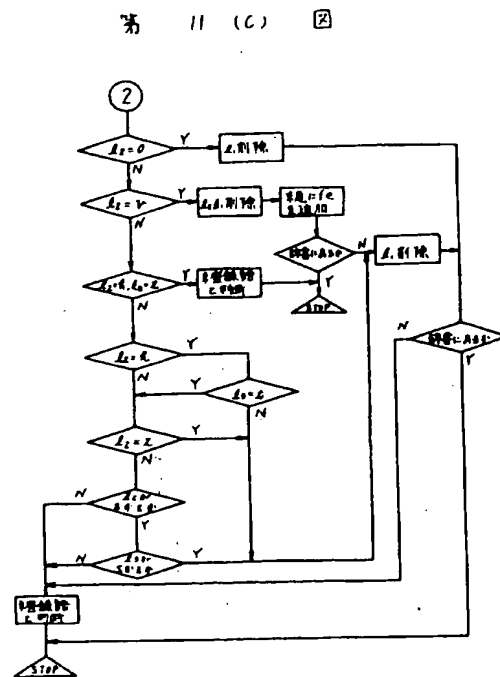
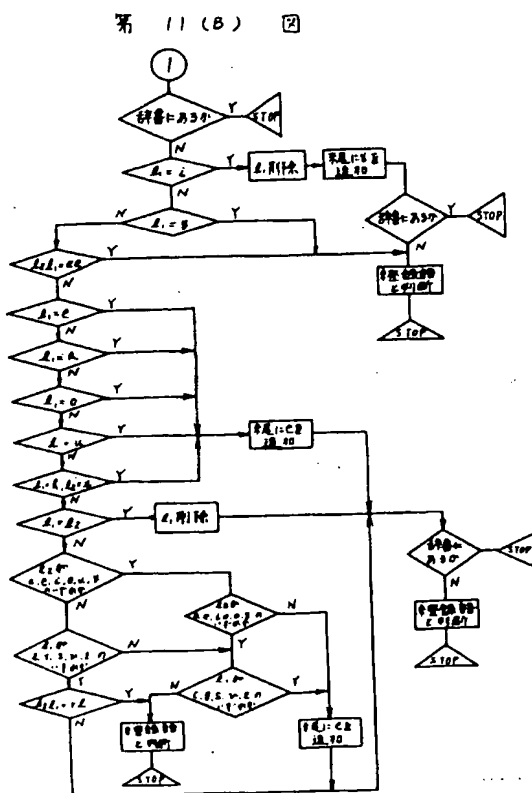
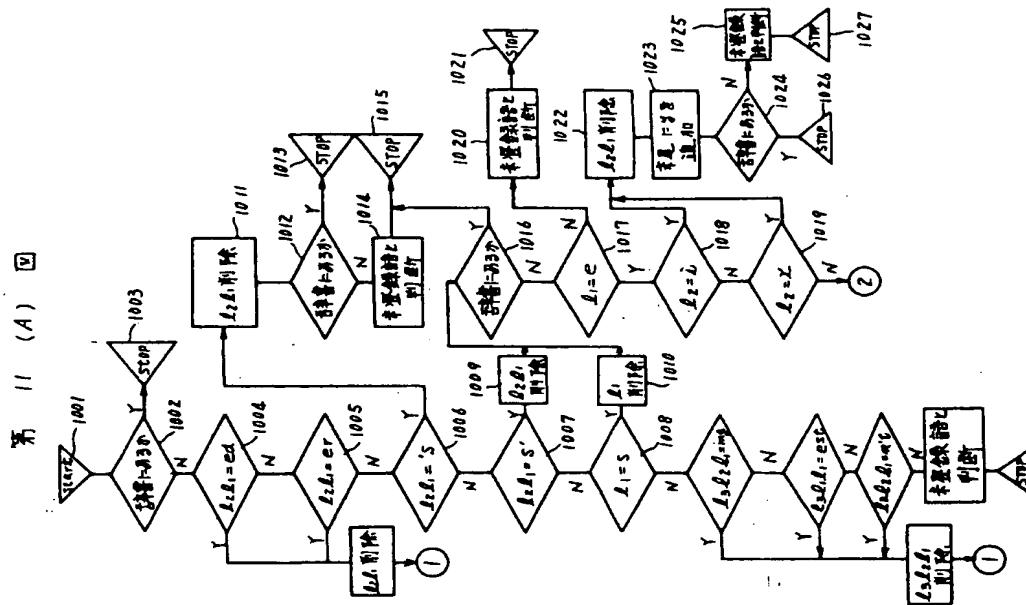
項番	断片上の並列	骨格ハローン
1	中 C 中	SENT ^S
2	SENT ^S DEL C AND BUT SENT ^S DEL	COMPD SENT ^S

第 9 図

1	2	3	4	5	6	7	8	9	10	11
H	E		W	R	O	T	E		E	N
12	13	14	15	16	17	18	19	20	21	22
G	L	I	S	H		V	E	R	Y	
23	24	25	26	27	28	29	30	31	32	33
S	L	O	W	L	Y			B	U	T
34	35	36	37	38	39	40	41	42	43	44
H	E		L	I	K	E	D			T
45	46	47	48	49	50	51	52	53	54	55
O	S	T	U	D	Y		T	O		
56	57	58	59	60	61	62	63	64	65	66
R	E	A	D		A	N	D		W	R
67	68	69	70	71	72	73	74	75	76	77
I	T	E		I	N		E	N	G	L
78	79	80	81	82	83	84	85	86	87	88
I	S	H		A	T		S	C	H	O
89	90	91	92							
O	L		(END)							

第 10 図

401	HE	WRITE	ENGLISH	
	単語 情報	単語 情報	単語 情報	
	WB(1)	WB(2)	WB(3)	WB(N)

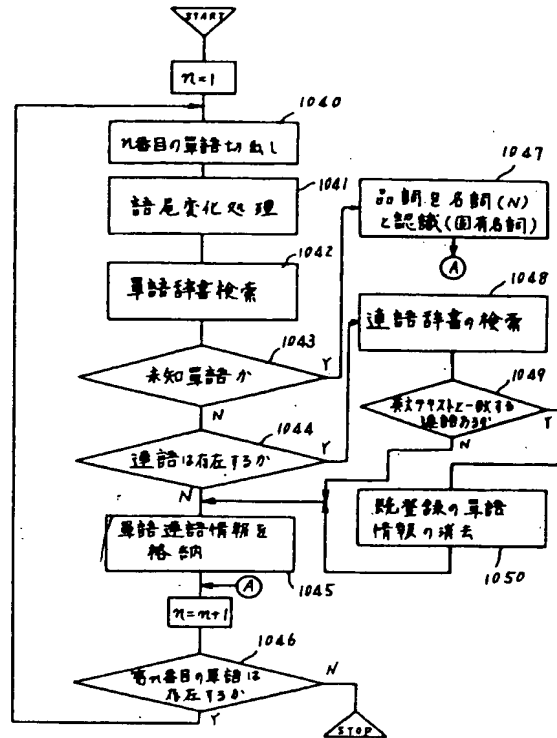


第 11 (D) 図

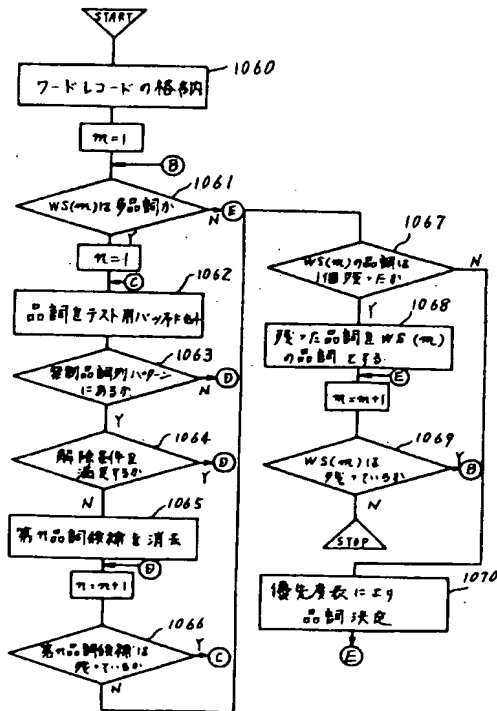
S	T	U	D	I	E	S		
L ₇	L ₆	L ₅	L ₄	L ₃	L ₂	L ₁		

第 11 (E) 図

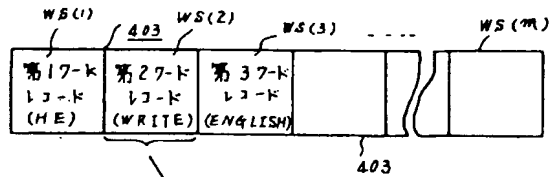
S	T	U	D	I		
L ₅	L ₄	L ₃	L ₂	L ₁		



第 13 図



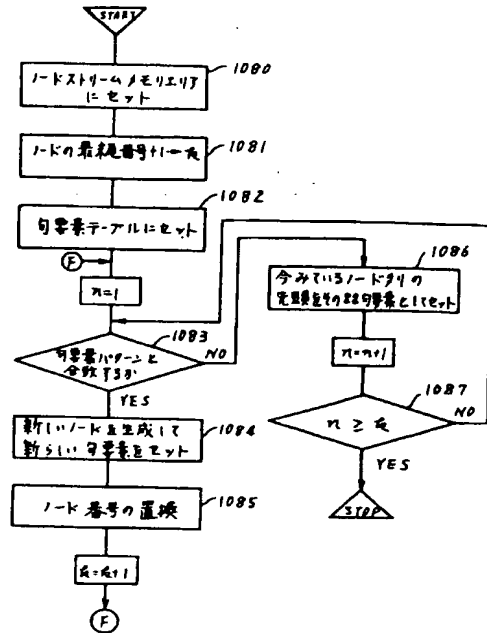
第 14 (A) 図



第 14 (B) 図

単語/連語の区別指示子 (単)	
語の識別番号 (2)	
語長 (5)	
単語連語の先頭文字の番号 (4)	
多品詞の個数 (2)	
第1候補品詞 (V)	第1候補品詞の分類
多義性の個数	第1品詞の先頭語の番号
第2候補品詞 (N)	

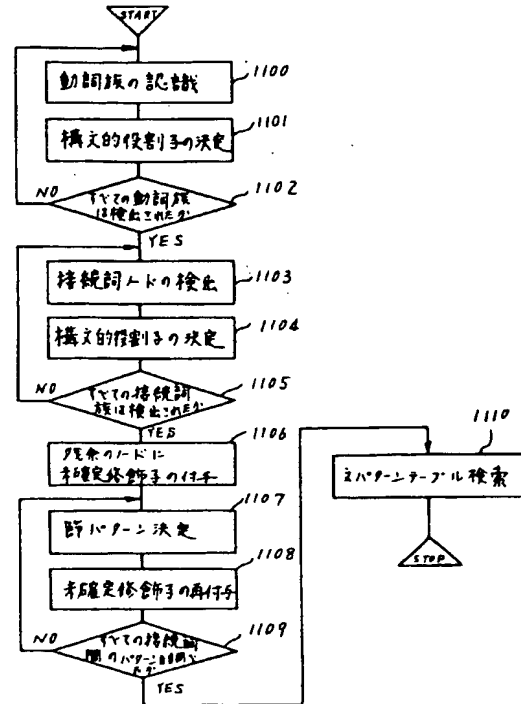
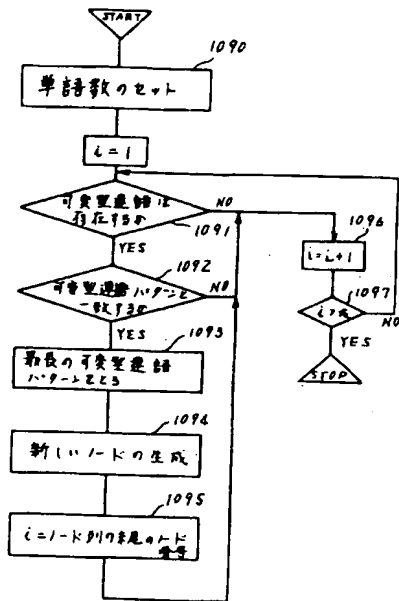
PRN													
N	=												
ADJ	>	>											
CNJ	>	>	<										
V		<	>										
AUX		>				>							
ADV	<	<	<	>	>								
PREP													
ART	>	>	>	>	<				>				
INT		>											
NOT		<									>		
	PRN	N	ADJ	CNJ	V	AUX	ADV	PREP	ART	INT	NOT		



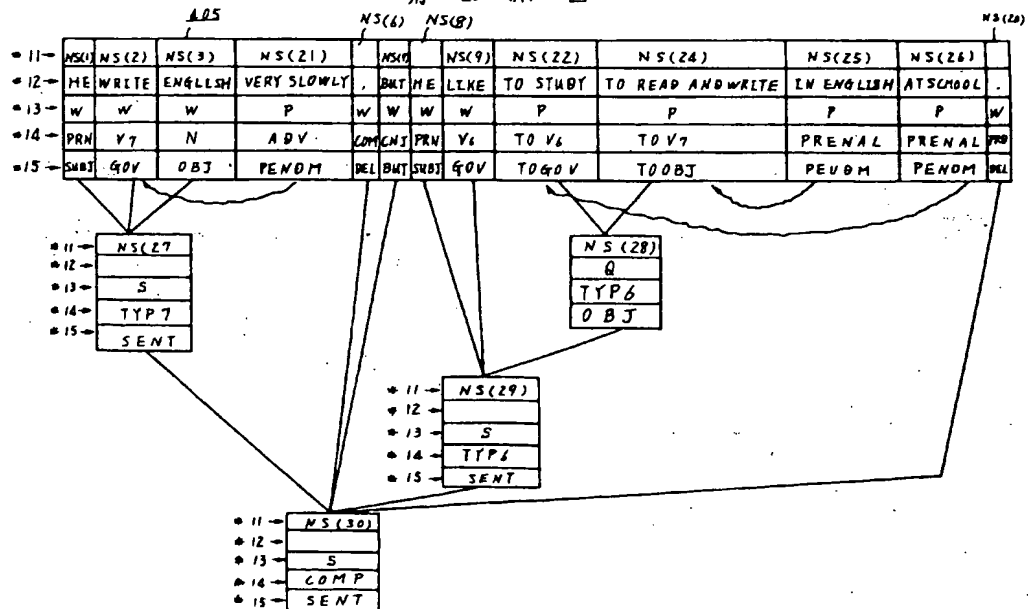
The diagram shows a sentence: "HE WRITE ENGLISH VERY SLOWLY, BUT HE LIKE TO STUDY TO READ AND WRITE IN ENGLISH AT SCHOOL". The sentence is segmented into tokens, each labeled with a number in parentheses (e.g., NS(1), NS(2), etc.). Below the sentence, a series of boxes represent the POS tags for each token. The boxes are labeled with the token number and the POS tag. For example, the first box is labeled "NS(1)" and contains the tag "P". The second box is labeled "NS(2)" and contains the tag "V". The third box is labeled "NS(3)" and contains the tag "N". The fourth box is labeled "NS(4)" and contains the tag "ADV". The fifth box is labeled "NS(5)" and contains the tag "ADV". The sixth box is labeled "NS(6)" and contains the tag "COM". The seventh box is labeled "NS(7)" and contains the tag "CONJ". The eighth box is labeled "NS(8)" and contains the tag "PRN". The ninth box is labeled "NS(9)" and contains the tag "V". The tenth box is labeled "NS(10)" and contains the tag "PREP". The eleventh box is labeled "NS(11)" and contains the tag "V". The twelfth box is labeled "NS(12)" and contains the tag "PREP". The thirteenth box is labeled "NS(13)" and contains the tag "V". The fourteenth box is labeled "NS(14)" and contains the tag "CONJ". The fifteenth box is labeled "NS(15)" and contains the tag "V". The sixteenth box is labeled "NS(16)" and contains the tag "PREP". The seventeenth box is labeled "NS(17)" and contains the tag "N". The eighteenth box is labeled "NS(18)" and contains the tag "PREP". The nineteenth box is labeled "NS(19)" and contains the tag "N". The twentieth box is labeled "NS(20)" and contains the tag "PRD".

第 18 回

第 19 回

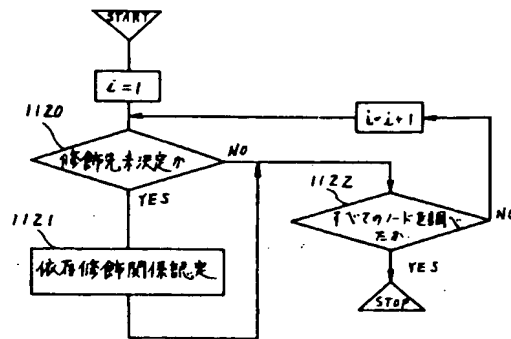
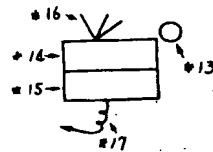


第 20 (A) 回

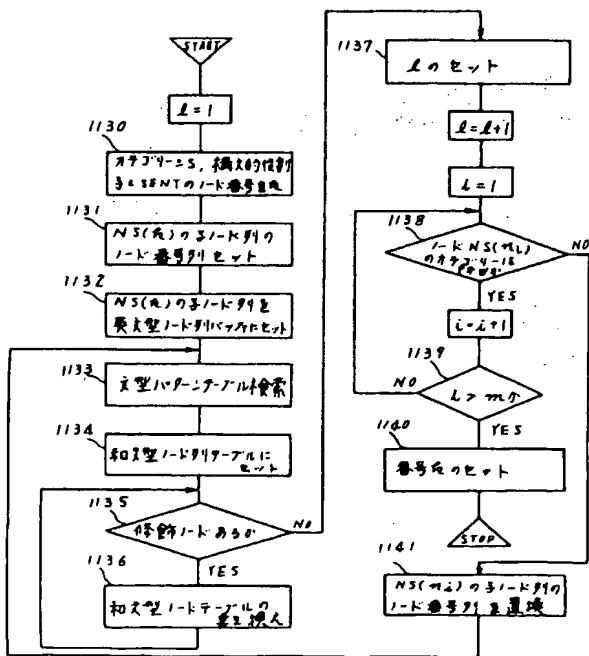


第 21 回

第 20 (B) 圖



22



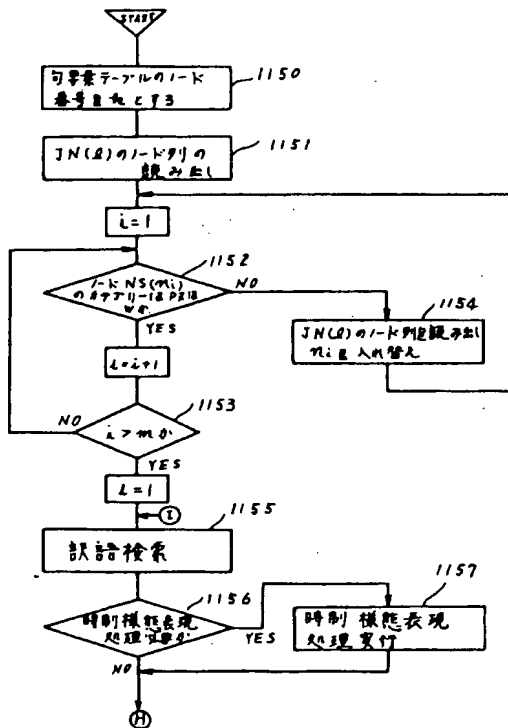
第 23 回

	EN(1)	EN(2)	EN(3)	EN(4)	EN(5)	EN(6)
	27	6	7	29	20	
	(SENT)	(DEL)	(BUT)	(SENT)	(DEL)	

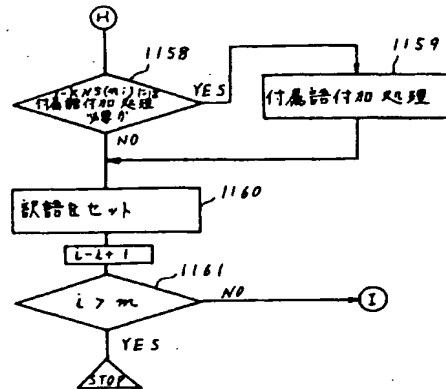
406

	(29)	(6)	(7)	(29)	(20)	
JN(1)	SENT	,	BUT	SENT	.	
	(1)	(3)	(20)	(2)		
JN(2)	S U B J	O B J	A D V A L	G O V		
	(9)	(25)	(24)	(24)	(22)	(9)
JN(3)	S U B J	A D V A L	O B J	A D V A L	O B J	G O V
JN(4)						

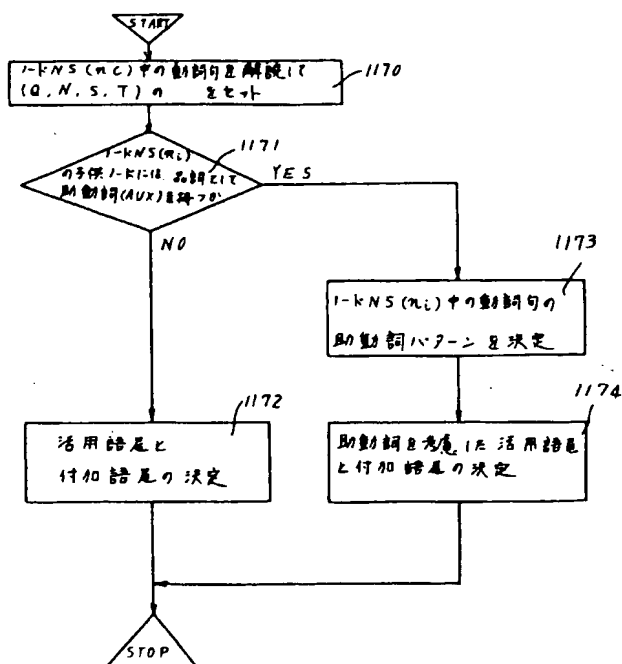
407



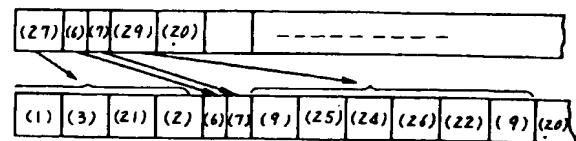
第 24 (B) 図



第 24 (C) 図



第 25 図



第 26 回

